UNITED NATIONS







Distr. LIMITED E/ESCWA/STAT/2001/WG.4/7 18 October 2001 ORIGINAL: ENGLISH

Economic and Social Commission for Western Asia

Workshop on Industrial Statistics Cairo, 4-8 November 2001

LIBRARY & DOLL SAMPON

UNIDO'S VIEW OF INDUSTRIAL STATISTICS AND TREATMENT OF NATIONAL OFFICIAL DATA

by

Tetsuo Yamada

Note: This document has been reproduced in the form in which it was received, without formal editing. The opinions expressed are those of the author and do not necessarily reflect the views of ESCWA.

		•
		,

UNIDO'S VIEW OF INDUSTRIAL STATISTICS AND TREATMENT OF NATIONAL OFFICIAL DATA

Tetsuo Yamada
Senior Industrial Development Officer
Statistics and Information Networks Branch
Investment Promotion and Institutional Capacity-Building Division
United Nations Industrial Development Organization (UNIDO)

This paper overviews (i) the relevance of industrial statistics in general and that of international industrial statistics in particular for the industrial development in the macro economic framework, (ii) the importance of data comparability on those statistics for accurate and objective analysis with regard to the industrial development, and (iii) UNIDO's efforts in increasing cross-country data comparability. Because of the interrelated topics, this paper and the previous paper, the *Structure and maintenance of the UNIDO Industrial Statistics Databases System*, are complementary to each other with some overlaps.

RELEVANCE OF INDUSTRIAL STATISTICS

- Information (together with theories and concepts) is essential for any decision-making. Because of the accelerating dynamics in socio-economic development together with increasing complexity in Information Technology system, quick, decisive and relevant decision making on solid factual grounds gains increasingly important.
- Statistical information is of prime importance because of its quantitative nature and the resulting features of objectivity and comparability. Only numerical data provide objective and comparable (and, thus, reliable) measures. Meaningful monitoring, interpretation and prediction of structure and performance relating to demand, supply, trade, investment, etc. can be made only empirically.
- In the era of globalization of economy, monitoring of sustainable industrial development and related economic-growth empirics require specialized and detailed economic industrial statistics as an indispensable information basis. Viable strategies, policies and programmes for sustainable industrial development and investment cannot be formulated unless they are well prepared based on such statistics and analysis. Those that are not based on relevant empirical evidence would be subject to dispute.

However, this argument holds <u>only if</u> the statistical data are qualitatively **comparable** within each of the specified data dimensions (e.g., countries, industries, years, etc.) and, needless to say, they are relevant and accurate.

Particularly in the era of the rapid globalization of industry, specialized and detailed information on industry is needed more than ever in order to assess implications of the process of the globalization for individual countries. Hence, a worldwide data set of internationally comparable national industrial statistics should be available to all researchers, observers, policy makers, potential investor, business people, project personnel and the like who are dealing with industry.

UNIDO deals with industrial development aiming ultimately at worldwide poverty alleviation through economic growth. The most recent macro-economic growth empirics (or, the new vision for such research) call for analysis at the sectoral and the sub-sectoral levels. Worldwide industrial statistics are to serve as the fundamental information input to such analysis, either cross-country or even country-specific.

Taking this into consideration and following decided international division of labour in compilation of statistical information among international agencies, UNIDO compiles detailed sub-sectoral data relating to **industrial production**. The organization maintains

international databases of selected set of industrial statistics. On the other hand, by employing the databases it carries out empirical economic research, for instance, on

- manufacturing production and its growth, location and pattern,
- international trade in relation to manufacturing production and comparative advantage,
- production factors of manufacturing (e.g., labour, human capital, fixed capital),
- manufacturing productivity,
- structure (e.g., value-added structure, trade structure, employment structure, market structure) of the manufacturing sector, and
- technology as a factor of manufacturing production.

UNIDO INDUSTRIAL STATISTICS DATABASES

The industrial statistics databases that are currently maintained by UNIDO are

- UNIDO Industrial Statistics at the 3-digit Level of ISIC(Rev. 2) (INDSTAT3)
- INIDO Industrial Statistics at the 4-digit Level of ISIC(Rev. 2) (INDSTAT4), and
- UNIDO Industrial Statistics at the 3- and 4-digit Levels of ISIC(Rev. 3) (INDSTAT-Rev3)

In addition to those three INDSTAT databases, UNIDO also maintains, within its industrial statistics (INDSTAT) database system, two other databases: the Industrial Demand-Supply Balance Database (IDSB) at the 4-digit level of ISIC(Rev.2) and the Industrial Statistics Size-Distribution Database at the 3-digit level of ISIC(Rev.2). Despite their usefulness in the analysis on, respectively, trade dependency and market structure of industry, those two databases are excluded from the scope of the current paper since IDSB, a production-cum-trade statistics database, is a derived database from INDSTAT4 and the UN Commodity Trade Statistics Database, and the size-distribution database is treated outside the INDSTAT system.

Formulation of a programme to secure food supply, for instance, would require quantitative data (i.e., data in physical measuring units) regarding arable land, cultivated areas, demand-supply balance of food products, etc.. Similarly, for the industrial

development aiming at economic development/growth in any country/region, a set of key economic industrial statistics is required as the basic statistical information source.

The data items in the three INDSTAT databases are limited to those for which majority of developing countries compile their national data on the regular basis. Furthermore, the databases focus on a minimum range of economic industrial statistics from the viewpoint of reduced reporting burden on national statistical offices (NSOs). Consequently, these databases refer to annual time-series on the following key industrial statistics for individual countries and areas:

- Number of establishments (EST)
- Total employment (EMP)
- Female employment (FEMP)
- Wages and salaries paid to employees* (SAL)
- Output* (OUT)
- Value added* (VAD)
- Gross fixed capital formation* (GFCF)
- Production indexes (IND) (INDSTAT3 only)
 - * in current national currencies.

Despite the limited number of statistics covered, the overall data maintained in the UNIDO's INDSTAT database system, together with the data of the United Nations Commodity Trade Database, are detailed enough and can be considered to be sufficient for basic cross-country cross-industry analysis on industrial development and related structural changes of the manufacturing sector such as those subjects listed in the previous section.

For consistent cross-country analysis on industry, cross-country comparable data on those statistics are required. Thus, thorough quality check and control of official data reported by countries are needed in terms of international comparability, which is described below in detail.

TREATMENT OF DATA REPORTED BY NATIONAL STATISTICAL OFFICES

In the context of globalization, international statistics for many social, economic and environmental areas that are collected, maintained and disseminated by international data sources have been increasingly demanded not only for cross-country analysis but also for country-specific analysis.

UNIDO requests individual country to provide its available national data on the selected industrial statistics in accordance with either Revision 2 or Revision 3 of ISIC system,

It should be noted here that commodity-level statistics (e.g., for production, consumption, trade) in physical units are limited in terms of commodity coverage. Furthermore, they are not really economic statistics and, thus, of limited validity in decision making for viable economic development and investment.

depending on the country's current national industrial classification scheme in use². Depending on the revision level of ISIC with which the reported data are arranged, data are stored in one of the three INDSTAT databases, INDSTAT3, INDSTAT4 and INDSTAT-Rev3, run in parallel. Data conforming to ISIC(Rev.3) are entered in INDSTAT-Rev3 and at the same time, whenever possible, converted to Revision 2 compliant structures, that is conforming to the 3-digit level of Revision 2, and entered in INDSTAT3 as well. This is done mainly in favor of maintaining updated historical time series ranging back as far as 1963.

The industrial statistics under the current consideration are highly detailed in terms of industry classification. Furthermore, they are measured/valued according to varying concept and definition. Thus, their information value is high but, at the same time, their reported data tend to be incomplete and inconsistent in terms of comparability across countries and sometimes even across variables, industries and over time because of possible difference in employed national standards and their changes. Therefore, validity of each reported datum must be judged carefully on the basis of international standards.

1. Potential sources of cross-country incomparability of data

Quality of statistical data has several dimensions. They include, for instance, relevance (for users' purposes), accuracy, timeliness, accessibility, comparability, coherence and completeness. UNIDO's quality assurance work on reported national data for those quality dimensions refers to, whenever possible; (i) identification and documentation of the sources of existing deviation from international standards and recommendations based on the respective metadata given by NSOs, (ii) adjustment of incomparable/incoherent data by employing available supplementary information, (iii) imputation and estimation of missing data by use of available supplementary information and econometric methods, and (iv) improvement of the organization's data publication and dissemination services together with the development of user-friendly data manipulation software incorporated in its data dissemination products.

Major determinants for the extent of international comparability of reported data are data coverage (or scope of the national establishment survey), employed concept and definition of the variable in question, and the national industrial classification, which are discussed below in some detail.

Data coverage:

Often reported data are known to exclude a significant portion of industrial activity, either because the coverage of small-scale establishments may be incomplete (e.g.,

² Presently, approximately two-third of some 120 regularly-reporting countries and areas are in a position to report their data in Revision 3.

exclusion of the establishments below a certain survey cut-off-point in terms of their size), the data may refer only to a certain area of the county (e.g., urban area, major provinces) or they may refer to only a part of the manufacturing sector (e.g., formal sector, registered establishments, selected industries, etc.). This characteristic is certainly the most challenging of all sources of data incomparability because adjusting for coverage involves the attempt to quantify what is not there. The problem of data coverage may be broken down into three parts: (i) incomplete or varying degrees of coverage of establishments; (ii) non-reporting of data by surveyed establishments, and (iii) the failure to adjust for non-response.

While the question of the treatment of non-response is basic for the data user, it has not received the attention that it deserves among many national data producers. Some countries adjust their data for non-response, and others do not. The latter usually provide some crude measure of the extent of non-response, which is used by the user to assess the quality of the data. However, some countries fail to address the question altogether. The International Recommendations³ specifically request such information, and perhaps this is an area where improvements in national reporting practices may be anticipated.

Concepts and definitions:

Differences in concept or definition are variable-specific although their numerical effects may vary across industries (i.e., ISIC categories). In reporting their industrial data, most countries conform to the United Nations' recommendations. Even among those countries that do not, the international standards provide a convenient reference point for comparing all variations in national reporting practices.

Employment: For the majority of the countries, employment data refer to number of employees. However, in some cases data refer to number of persons engaged. For a few countries, the definition changes over time. In general, no supplementary information for standardization of reported employment data is available. Any use of employment data, therefore, requires caution, particularly in those cases where definition changes over time.

Wages and salaries: In the reporting of wages and salaries, the most common differences between national practices and the international recommendations relate to the inclusion of payments to family workers and of employers' contributions to social security schemes or the exclusion of payments-in-kind. The numerical effects of these differences, although not known, are probably of small consequence both within and between countries, compared to the effects of differences in survey cut-off point.

Output and value added: Among the variations in concept that may apply to the data on output and value added, the most important (in terms of measuring consequence to data) are: (i) the difference between the national accounting concept and the industrial census

³ International Recommendations for Industrial Statistics, Statistical Papers, Series M, No.48, Rev.1, United Nations, 1983, paragraph 7.

concept; and (ii) the difference between valuation at producers' prices, valuation in factor values and other valuations. The main difference between the national accounting concept and the industrial census concept is in the treatment of non-industrial services. On the other hand that in the different valuations is in the treatment of indirect taxes and subsidies. These differences can be significant, and should be taken into account especially if comparisons between countries are being made.

2. Recent Data Situations of the ESCWA Member Countries

In view of the international comparability of data on the industrial statistics under the current consideration, the existing drawbacks with regard to the data reported to UNIDO by the NSOs of the ESCWA member countries are summarized below.

i. Only a limited number of the countries are in a position to report the requested industrial statistics: During the past four years, eight of the thirteen ESCWA member countries have reported available data through the UNIDO country questionnaire (see Table 1). Among these eight countries, only three countries (Egypt, Jordan, Kuwait) have reported data on all seven industrial statistics⁴.

Table 1. Data-reporting ESCWA countries in the recent rounds of the UNIDO INDSTAT annual compilation programme

UNIDO INDSTAT annual compilation programme (latest ref.year)	
	ESCWA member countries that reported data (Latest reference year)
1998 round (1996)	Egypt(1994), Jordan(1995), Kuwait(1995), Lebanon(1996), Oman(1995), Yemen(1996)
1000 1 (1007)	
1999 round (1997)	Egypt(1995), Jordan(1997), Kuwait(1996), Oman(1997), Syrian A.R.(1997)
2000 round (1998)	Egypt(1997), Iraq(1998), Kuwait(1997), Oman(1998), Syrian A.R.(1998)
2001 round*(1999) Kuwait(1998), Lebanon(1998), Oman(1999)	
* Ac of 20 Comba	

^{*} As of 30 September 2001.

ii. Among them, only a few are in a position to report the requested data with a two-year time lag: The latest reference year of the 2000 round of the UNIDO INDSTAT annual compilation programme was 1998. However, by the end of 2000, only three countries reported some/all of the requested data for 1998. The

⁴ Collection of index numbers of industrial production (IND) is under the responsibility of the Statistics Division of the United Nations (UNSD). To complete the UNIDO INDSTAT Databases, UNSD forwards the data on IND, that were reported to UNSD by NSOs, to UNIDO. UNIDO processes the reported indexes in accordance with its INDSTAT-Databases' requirements. Examples include transformation of the reported indexes to series with a common base year and aggregation of them to derive those at the 3-digit level of ISIC(Rev.2) or 2-digit level of ISIC(Rev.3) by applying value-added weights .

reported data have been often incomplete in terms of variable coverage as well as industry coverage. Furthermore, due to incompleteness of the provided metadata, the consistency of the reported statistical data is often difficult to be judged. With respect to the years 1990 onward, the current period coverage and existing drawbacks of the national data reported by the NSOs of the ESCWA countries are summarized in Table 2. By now, five of the thirteen countries have started reporting their data in accordance with Revision 3 of ISIC.

In the table, the following drawbacks regarding the reported data for the ESCWA countries are, except for a few countries, generally observed:

- Incomplete coverage of the variables.
- Incomplete time series and data inconsistency over time.
- Lack of relevant metadata that indicate the comparability and quality of the reported statistical data.
- Cross-country, cross-variable and/or cross-year incomparability due to different/changing combinations of ISIC categories, possible no adjustment of non-response, different definitions of variables, etc.

All these problems limit the usefulness of the data in cross-country comparative analysis, country-specific trend analysis, etc. of industrial growth and structure.

Table 2. Availability and quality of the data for 1990 onward reported by the ESCWA countries

Country	Latest ref. year (as of 30 Sept. 2001)	ISIC disaggregation level	J J J J J J J J J J J J J J J J J J J	Unavailable basic metadata concerning comparability and quality of reported statistical data
	1992	Rev.2 4-dig.	VAD missing; IND available only for one 3-dig. branch; No time series (data reported for 1992 only)	survey scope; treatment of non- response; variable definitions
Egypt	1997	For 1997, Rev.3 4-dig.; For years up to 1995, Rev.2 4- dig.		survey scope; treatment of non- response
Iraq	1998	For 1991 and 1992, Rev.2 3- dig.; For 1995- 1998, Rev.3 4- dig.	GFCF and IND missing; Data available only for scattered years; Rev.3 data were reported only for 1998 except for VAD for which incomparable data were reported for 1995-1997. Rev.3 data were reported only for a few ISIC categories.	survey scope; treatment of non- response; variable definitions
Jordan	1997	For 1994 onward, Rev.3 4-dig.	Data on IND for several 2-digit categories missing	
Kuwait	1998	Rev.2 4-dig.	IND reported only for one 3-dig. branch)	treatment of non- response; variable definitions
Lebanon	1998	For 1996 onward, Rev.3 4-dig.	For year up to 1995, no data reported; For 1996, only EST and EMP reported; For 1998, data were reported through a hardcopy publication.	survey scope; treatment of non- response; variable definitions
Oman	1999	From 1993, Rev.3 4-dig.	FEMP and IND missing; Due to different survey scope, 1999 data are not comparable with data for other years.	variable definitions
Qatar	1994	Rev.2 3-dig.	FEMP, GFCF and IND missing; Several branches are combined.	survey scope; variable definitions
Saudi Arabia	1992	Rev.2 2-dig.	Only EST and EMP reported; No time series (data reported for 1992 only);	variable definitions
Syrian Arab Rep.	1998	Rev.2 2-dig.	EST, FEMP and GFCF missing	treatment of non- response
UAE	n.a.	n.a.	No data reported	
Yemen	1996	Rev.2 4-dig.	Up to 1994, only EST, EMP and OUT reported at Rev.2 2-digit level with some combinations; For 1995 onward, only EST and EMP reported at Rev.2 4-digit level; IND reported for only	treatment of non- response; variable definitions
Palestine	n 0		Rev.2 3-dig. category	
arestine	n.a.	n.a.	No data reported	

- iii. In principle, improvement of data quality can be done only by the responsible NSOs. NSOs' possible solutions to the drawbacks of respective future national data include:
 - a) Regular implementation of annual industrial surveys following a common (e.g., the UN-recommended) methodology and survey scope (e.g., employment cut-off point of establishment).
 - b) Continuous update of the directory of establishments.
 - c) Periodical implementation of complete industrial censuses to obtain a complete picture of industry in terms of key industrial statistics. The results of annual surveys are to be compared with census results. Industrial censuses covering all establishments regardless of their size can produce detailed data on industrial statistics by size of establishment. Size-distribution data, which themselves are very useful information on market structure of industry as well as on the SMEs, can be a good benchmark to assess annual survey data.
 - d) Adoption of ISIC, particularly of its Revision 3, or a national industrial classification that is compatible with ISIC Rev.3.
 - e) Adoption of the UN-recommended concepts and definitions of industrial statistics.
 - f) Maintenance of data consistency over time and across variables.
 - g) To ensure proper use of the statistical data, preparation and dissemination of user-friendly metadata that indicate the quality of the statistical data from the viewpoint of deviations from international standard methodology, survey scope, concepts and definitions of variables and industrial classification (e.g., ISIC).
 - h) For many of the 13 countries, production indexes (IND) at the given level of aggregation are either incomplete or totally missing. Since no deflators for manufacturing output at any subsectoral level are available, IND are indispensable to assess the real growth of industrial production at a subsectoral level of manufacturing. Thus, efforts should be made to produce reliable production indexes at a disaggregated level (e.g., 3-digit level of ISIC).
 - i) With regard to gender statistics, it is realistic to assume that data concerning only employment and wage bill can be compiled. However, according to UNIDO's investigation, many countries worldwide actually do not compile/publish data on wage bill by sex due to difficulties in obtain such information from individual establishments.
 - j) Increase in the efficiency of the NSOs in the all phases of their data compilation work including survey organization and management, field enumeration, estimation, editing, tabulation and publication of data to realize timely dissemination of reliable and comparable data. Here, effective quality assurance at every work phase is crucially important.

3. Ensuring Data Consistency

To improve the international usefulness of the data reported by NSOs, not only the correction of simple reporting errors based on the NSOs' responses to concerned data queries by UNIDO but also adjustment of the reported data to achieve international standards as well as estimation and imputation of missing data to increase data coverage are required. For UNIDO, these tasks are possible only if appropriate supplementary statistical information is available. Furthermore, they require sector-specific statistical as well as economic knowledge. As for international industrial statistics, UNIDO's statistical unit, which has not only statistical but also empirical-economic-research capabilities in the field of global industry, is in a strong position to carry out such data generation⁵.

UNIDO statisticians take considerable care in ensuring data consistency in the process of enlarging the INDSTAT databases and in improving the international comparability of their contents. However, due to inconsistency inherent to many series reported by primary and other sources, it is felt that a final screening of the data is needed. The purpose of this final screening is to diagnose and display 'abnormal' entries in the databases, to allow for possible corrections. The final screening takes place in two phases. First, possible abnormalities are identified through a computerized procedure. Second, UNIDO statisticians redress, to the extent possible, the identified abnormalities.

UNIDO's procedures to treat the national data reported by each NSO are summarized as the following:

- a) Pre-filling of the out-going *UNIDO General Industrial Statistics Questionnaire* with previously reported statistical and meta data for their possible revision by the NSO.
- b) Upon receipt of the questionnaire completed and returned to UNIDO by the NSO, (i) manual detection and, if possible, correction of incoherent data (e.g., mismatch between sum of components or disaggregated data and given total) and dubious data (e.g., obvious typo errors, abnormally fluctuating data, those resulting abnormal relations between variables), (ii) adjustment of reported statistics to desired statistics (e.g., if data on wage rate are reported instead of wage bill, deriving data on wage bill from reported data on per employee wage rate and those on the number of employees), and (iii) if appropriate, re-description of the provided metadata from the viewpoint of international comparability.

⁵ The UNIDO's capabilities create a strong position also to provide related statistical services to users, generate demands for statistical information on industry, promote appropriate use of the information through its consultancy services to users and to monitor changing data demands.

- c) If the validity of the dubious data cannot be judged, enquiry to the NSO for clarification of them, and correction of them in accordance with the clarification made by the NSO.
- d) Incorporation of the corrected and adjusted data in a relevant INDSTAT database. If the reported data are in Revision 3 of ISIC, incorporation of the data in INDSTAT-Rev3 database and, at the same time, conversion of them to those at the 3-digit level of ISIC Revision 2 and incorporation of the converted data in INDSTAT3 database. Available metadata are incorporated in a metadata file common to all INDSTAT databases.
- e) Computer-based screening (consistency and coherency check) of the initially updated INDSTAT databases to further detect dubious data, and manual correction/rejection of the detected data (if deemed necessary, clarification is requested to the concerned NSO.).
- f) Computer-based estimation of missing data by estimating inter-variable relations as country-specific time-series regression equations (Such estimation is attempted only for four variables, namely total employment, wage bill, output and value added).

4. Development of User-Oriented Metadata in Support of INDSTAT Databases

Cross-country incomparability, in many cases, of reported data is due to different national needs and practices in industrial statistics for legal, economic and policy purposes. However, adjustment of incomparable data in line with international comparability as well as estimation of missing data is often difficult to be made. Thus, the coverage and the extent of the comparability of the available data in each INDSTAT database are limited. To assure sound use of those data, UNIDO, as in the case of any other supplier of statistical information, is obligated to provide users with, in addition to general information for the overall database, relevant and detailed qualitative information (e.g., information concerning the deviations from international standards or norms) that indicates the applicability/limitation of individual disseminated statistical figures in terms of international comparability.

Preparation of appropriate statistical background information (metadata) in support of INDSTAT databases requires concrete and well-documented metadata inputs from primary data compilers. Thus, UNIDO requests NSOs to provide, together with available statistical data, such descriptive information through its industrial statistics country questionnaire. The key items for which the organization needs to obtain meta information include:

- Name of the supplier of the statistical data (i.e., the reporting NSO),
- Basic source of data (e.g., annual industrial survey),
- Major deviations from ISIC,
- Reference period (e.g., calendar year),

- Scope of the annual survey,
- Method of enumeration and data compilation,
- Treatment of non-response,
- Concepts and definitions of variables,
- Related national statistical publications, and
- Cell-level footnotes, if any.

The metadata that are provided by NSOs often do not explicitly indicate deviations from international standards. In such cases, UNIDO attempts to re-describes/re-arranges the provided metadata into explicit information concerning the deviations from the international standards. This is often a difficult task and requires additional clarifications from the concerned NSO.

Countries have been increasingly introducing international statistical standards because of the reason stated earlier, which has resulted in less needs of metadata on international statistics and in more relevant metadata inputs from NSOs from the viewpoint of international standards in recent years.

5. Concluding Remarks

As briefly stated above, UNIDO makes efforts to adjust the official data reported by NSOs to increase the international comparability of those data and to estimate missing observations to increase the data availability of its databases. The data adjustments are made, as much as possible, in a way that the adjusted data would be coherent with the corresponding reported official data in order to avoid the situations of conflicting data between the two data sources and consequent confusion among users. Furthermore, adjusted data and estimates generated by UNIDO are footnoted as UNIDO estimates. With regard to INDSTAT3 database, for instance, approximately 30 per cent of its contained statistical data are those generated (i.e., adjusted or estimated) by UNIDO.

Achievement of the complete data coverage and, in particular, that of the cross-country comparability of data by employing common standards across countries are the top priorities in the maintenance of the INDSTAT database system. However, such achievement by UNIDO alone is limited without consensus in definition, adoption and implementation of international standards in the international statistical community involving all NSOs.

For this, UNIDO plays a significant role, under the initiative of the United Nations Statistics Division, in the international statistical community for the coordination, development and promotion of international statistical classifications, concepts and definitions pertaining to industrial statistics, aiming at:

- Increased relevance of industrial statistics reflecting the rapidly changing nature of industrial activities,
- NSOs= adoption of international standards and concepts, and

 Metadata development by NSOs in support of their industrial statistics from the viewpoint of good practice of statistical services.