OCCASION

This publication has been made available to the public on the occasion of the 50th anniversary of the United Nations Industrial Development Organisation.

DISCLAIMER

FAIR USE POLICY

CONTACT

Please contact publications@unido.org for further information concerning UNIDO publications.

For more information about UNIDO, please visit us at www.unido.org

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION
Vienna International Centre, P.O. Box 300, 1400 Vienna, Austria
Tel: (+43-1) 26026-0 · www.unido.org · unido@unido.org

Compilation of Index of Industrial Production in Developing Countries: Tanzania's Experience

UNITED NATIONS
INDUSTRIAL DEVELOPMENT ORGANIZATION

Acknowledgements

This report was prepared by Dong GUO, UNIDO staff member, and Willem van den Andel, international expert on industrial statistics, under the overall guidance of Shyam Upadhyaya, the Chief Statistician of UNIDO.

The report is based on one of the main outcomes of the technical assistance projects of UNIDO to strengthen the statistical capacity in the United Republic of Tanzania under the One UN funding support, entitled "Wealth Creation, Employment and Economic Empowerment" and "Country Framework of Support to UNDAP 2011-2011-2015". The report could not have been finished without the considerable support of the National Bureau of Statistics (NBS) in Tanzania during the period when the IIP for Tanzania was compiled. Mr. Child Hamisi and Ms. Veronica Claud, statisticians from NBS, Tanzania, provided valuable support and cooperated closely with international experts on IIP compilation. Ms. Joy Sawe and Mr. Fadhili Khalfani, Chief Statisticians at NBS, Tanzania, coordinated and supervised the activities, Mr. Morris Oyuke, Director of Economic Statistics at NBS, Tanzania, provided continuous support and guidance through the entire IIP compilation process. Last but not the least, Mr. Huzaifa Zoomkawala, international expert of system analysis, provided training to NBS, Tanzania, on the establishment of the IIP system and supervised the implementation system at NBS as well.

The authors would like to express their gratitude to Mr. Ludovico Alcorta, Director of the Development Policy, Research and Statistics branch, for his valuable comments and support. The authors are also grateful to Ms. Niki Rodousakis for editing the report. The authors are responsible for any remaining errors.

Table of Contents

Tables

List of Abbreviations

ASIP	Annual Survey of Industrial Production (Tanzania)
COICOP	Classification of Individual Consumption by Purpose
CPI	Consumer Price Index
CTI	Confederation of Tanzania Industries
CVA	Census Value Added
GDP	Gross Domestic Product
IIP	Index of Industrial Production
IMF	International Monetary Fund
ISIC	International Standard Industrial Classification of All Economic Activities
MITM	Ministry of Industry, Trade and Marketing
NBS	National Bureau of Statistics (Tanzania)
NPISH	Non-Profit Institutions Serving Households
PIMI	Production Index of Manufacturing Industries (the name of the IIP in Tanzania)
PPI	Producer Price Index
PRIMA	Price Index Management System
QNA	Quarterly National Accounts
SDDS	Special Data Dissemination Standards
SPD	Structured Product Description
TDV	Tanzania Development Vision
UN	United Nations
UNIDO	United Nations Industrial Development Organization
VA	Value Added
VAT	Value Added Tax

1. Introduction

The Index of Industrial Production (IIP, hereafter) is a very important indicator of economic performance. The IIP provides information for economic policymaking, which is used by stakeholders to gauge industrial performance and to forecast future economic performance. Many countries compile an IIP, as it is one of the infra-annual indicators of short-term economic development. When compiled at a high frequency (e.g. monthly), it can be used to identify the turning points in economic development at an early stage. The major advantage of the production index compared to other indicators is the combination of fast availability and detailed breakdown of activities. The theoretical aim of the IIP is to reflect the volume developments in value added (VA). Value added is defined as the difference between the value of gross output produced and the value of intermediate inputs consumed.

The IIP has been part of the statistical reporting system established by the United Nations since 1951, and has been included in its regular publications such as the Monthly Bulletin of Statistics. The vast majority of countries regularly compile an IIP and many report them to the United Nations for inclusion in its publications. As of September 2012, only 60 countries reported an IIP to the UN, which included 36 developed and 24 developing countries.

Among the countries that do not report an IIP, the compilation of the series is given low priority because of the countries' very low levels of industrial activity. This, for instance, applies to many small island nations. Nonetheless, there still is a substantial number of larger developing countries, mainly in Africa, which do not compile an IIP (or have discontinued doing so), often due to budgetary limitations. Reporting the data in a timely manner is crucial, as these are infra-annual data that lose much of their usefulness with the passing of time. Developing countries as a group have much bigger lags in reporting the information than industrialized countries.

The purpose of this publication is to provide an example of how a developing country can compile an IIP and disseminate it regularly. UNIDO Statistics has a technical assistance project in Tanzania[1] to build the country's statistical capacity in industrial

[1] The UNIDO project covers activities in both Tanzania mainland and Zanzibar. Administratively, those two entities are mostly separated and the discussions on this note refer to the mainland only. A similar IIP system has been developed for Zanzibar as well, but as it follows the same outline, it is not described here.

statistics. Within this project, it has, amongst others, supported the 2008 Annual Survey of Industrial Production (2008 ASIP, hereafter), which covers all larger industrial establishments in the country using an employment criterion. All establishments with 10 or more employees were included. The information was used to build a statistical database on industrial activities in the country. This database was subsequently linked to the existing dataset of the PIMI[2] to derive a new weights set and to update the list of companies.

The report is organized as follows. After a brief introduction of index concepts in the first section, Section 3 describes the experience of compiling Tanzania's IIP. The final section includes recommendations on IIP compilation for other developing countries. In the Annex, the integrated indexing software PRIMA is introduced.

2. The theoretical concept of IIP and compilation recommendations

2.1 Index number: a brief theoretical concept

An IIP describes the change of the volume of goods or services produced over time by measuring the changes of industrial production in value added, using constant prices. The general concept of production is defined in the 2008 System of National Accounts (SNA 2008). Production is carried out by institutional units[3] which can be corporations, government entities, unincorporated household enterprises or non-profit institutions serving households (NPISH). Production is the activity carried out under the responsibility, control and management of an institutional unit, which uses inputs of labour, capital, and goods and services to produce outputs of goods and services. Labour and capital (including land) are factors of production while the total inputs of goods and services into the production process are called intermediate consumption.

Value added is the balancing item in the production account. It is calculated as the difference between gross output and intermediate consumption. The main components of value added are labour costs and the remuneration of capital in all its

---

2 The PPI in Tanzania is called Production Index of Manufacturing Industries and has traditionally been limited to Manufacturing – Division C in revision 4 of the ISIC.

3 An institutional unit is an economic entity that is capable, in its own right, of owning assets, incurring liabilities and engaging in economic activities and in transactions with other entities.

forms. If defined as “gross” value added, the value of consumption of fixed assets is also included. When the valuation of gross output is in basic values, net taxes on products (taxes minus subsidies) are excluded from gross output and value added. These items are included when valuation is given at market prices. The four permutations of value added are given in Table 1. Note that gross output can also be measured in basic values and at market prices. Intermediate consumption is always given in purchasers' values, that is, in market values[4].

Table 1 Four permutations of value added

Value added definitions	Including consumption of fixed capital	Excluding consumption of fixed capital
Excluding net indirect taxes	**Gross** value added in **basic price**	**Net** value added in **basic price**
Including net indirect taxes	**Gross** value added at **market price**	**Net** value added at **market price**

Traditionally, another value added concept has been used in industrial statistics, usually described as Census Value Added (CVA). While the SNA value added concept is now recommended, most industrial surveys continue to provide information on CVA as well, even if not labelled as such. The difference between the two value added concepts lies in the partial exclusion of services from intermediate consumption in CVA. In industrial surveys, services are divided into industrial services, that is, those closely linked to the production process, and into non-industrial services. Under the first category, items such as repair and maintenance and subcontracted activities are included while the latter encompasses advertising, sales expenditures, etc. For analytical purposes, this difference is important as the consumption and production of industrial services is part of the production process while the use or production of non-industrial services is more or less discretionary and not linked to the production process in any way. Thus, while the production of cement, for instance, requires a certain amount of lime stone, gypsum, fuels and maintenance per tonne output, the decision to advertise depends on market conditions and other considerations. Branded products require heavy advertising in order to maintain market shares while own-label producers generally only need to advertise in trade journals and the like. In industrial analysis, the CVA concept is therefore considered important.

4 Market values/market prices can be measured in two ways, namely when received by the user (purchasers' prices) and when shipped by the seller (producers’ prices). The difference between the two is the trade and transport margins (TTM) incurred in the transaction.

Production is classified by activities, which are often also referred to as industries, as the title of the most important activity classification, the ISIC, indicates. The full title of this publication is "*International Standard Industrial Classification of All Economic Activities*". Traditionally, industrial statistics have been limited to extracting and converting activities in the economy, that is, to mining, manufacturing, electricity, water supply and construction, which in ISIC Rev.4 are divisions B to F. As the mandate of UNIDO does not include construction, statistics maintained by UNIDO cover the ISIC divisions B to E only. An important reason for defining the scope of industrial activities is to ensure internationally comparable IIP among countries. The activities covered in the recommendations are listed below, using the official terminology of ISIC rev.4.

B: Mining and quarrying

C: Manufacturing

D: Electricity, gas, steam and air conditioning supply

E: Water supply, sewerage, waste management and remediation activities.

Industrial production can be measured in values, either at current price or in volumes, that is, in constant prices. Thus, if P_t and Q_t represent price and quantity in the current period and P_0 and Q_0 represent price and quantity in another period, say, the reference period, then the value measure of industrial production is $P_t \times Q_t$ and the volume indicator is $P_0 \times Q_t$.

Index series are calculated at the basic level by comparing the value or volume of production in the current period with that in a reference period. If monetary values have been used to measure output, the current values need to first be deflated with an appropriate price index to obtain constant price values for the current period. The result of this first step is a set of volume ratios for the elementary items in the index series, which are without dimension and thus can be efficiently manipulated in a mathematical way. Traditionally, these ratios are expressed as percentages, thus giving a value of 100 for the reference period. These elementary indices are subsequently aggregated as weighted averages to obtain the indices at the different levels of detail for publication.

There are three main types of index numbers which are widely used internationally. These are the Laspeyres, Paasche and Fisher index numbers. The Laspeyres index

measures the volume changes at the prices of the reference period (Equation 1), while the Paasche index number values the changes at the prices of the current period (Equation 2). The Fisher index number is the geometric mean of the Laspeyres and Paasche indices for the same period (Equation 3).

$$L_t = \frac{\sum P_{i,0} Q_{i,t}}{\sum P_{i,0} Q_{i,0}} = \sum (W_{i,0}) \frac{Q_{i,t}}{Q_{i,0}} \quad (1)$$

$$P_t = \frac{\sum P_{i,t} Q_{i,t}}{\sum P_{i,t} Q_{i,0}} = \sum (W_{i,t}) \frac{Q_{i,0}}{Q_{i,t}} \quad (2)$$

$$F_t = \left[L_t \quad x \quad P_t \right]^{1/2} \quad (3)$$

Where:

$P_{i,o}$ = price for industry i at reference period 0;

$P_{i,t}$ = price for industry i at reference period t;

$Q_{i,o}$ = quantity for industry i at reference period 0;

$Q_{i,t}$ = quantity for industry i at period t;

$W_{i,o}$ = weight for industry i at reference period 0;

$W_{i,t}$ = weight for industry i at period t.

The base-weighted Laspeyres index is normally used for the production index series. If a producer price index series is maintained simultaneously, this series can also be prepared as a current-weighted Paasche index. To obtain a proper value series, the Fisher index can then be calculated as the geometric average of the Laspeyres production and Paasche price indices for the period.

2.2 Compilation guideline

The description in the previous section reveals that the IIP is, theoretically, well defined. When implementing an IIP, a number of practical issues need to be resolved. These include, for instance, how to measure industrial production in volume terms; what the data sources are; what the observation units are; which type of index will be used; how to cope with missing data; and how to ensure timeliness of the indices, as the IIP is used to estimate the quarterly GDP (QNA) for the industrial sectors in most countries that produce the QNA. International recommendations for all countries are

to produce QNAs in coming years, and that an IIP is necessary for all countries that have a sizeable industrial activity in terms of its share in GDP.

2.2.1 The measurement of industrial production in practice in terms of volume

The purpose of the IIP is to measure changes in the production volume of different activities measured in terms of value added over time. That is not always straightforward, as obtaining information on value added at a high frequency, for example, on a monthly or quarterly basis, is mostly not possible or, if possible, might take so much time that the resulting time lag is unacceptable. Therefore, value added is assumed to be a constant share of output and, with this constant input-output ratio by definition, also of intermediate inputs. Thus, the volume of output can be used as a proper measure of value added. In practice, there are two main approaches to measure industrial production for the IIP: (1) measurement of output, and (2) measurement of some inputs.

Output is measured in a number of different ways, depending on the type of industrial unit. Output indicators are:

(1) Value of output,
(2) Value of output sold, and
(3) Physical quantity of output.

When estimating industrial production from the input side, the relevant indicators are:

1. labour input, and
2. material consumed.

2.2.1.1 Measure of industrial output

The value of output includes all commodities produced, whether sold, otherwise used or entered into inventories for sale. The change in "work-in-progress" between the beginning and the end of the period is also included. All these outputs produced during the month or quarter should be included for the IIP. Valuation should be in basic prices, that is, excluding taxes and subsidies. For a proper comparison, all monetary values need to be expressed in prices from the same period. The IIP base/reference period is normally used as the pricing base.

Where the products are homogeneous, the physical quantity of output, which measures output in terms of the number of items, tonnes, litres, etc., can be used to track the development of production. The physical quantity of output approach to measure industrial production is the most suitable one for those industries that produce homogeneous goods where quality remains constant over time. This includes major industrial enterprises such as cement factories, oil refineries, flour mills, sugar factories, etc.

The term "value of output sold" is synonymous with turnover, sales or shipments and refers to goods or services sold by the given statistical unit during the reference period. The "value of output" described earlier does not include changes in "work-in-process", but includes the net increase in inventories of goods produced. For many companies, this information is easier to provide than data on actual production during the period. Note that there are some methodological differences when compiling IIP using "value of output" and "value of goods sold" that can affect the quality of the IIP. For seasonal products, which are produced in a limited period but sold throughout the year (for instance, sugar in most countries), the "value of goods sold" is an inappropriate indicator of production. For companies that produce to order and have little "work-in-process", both indicators provide similar results. For companies producing for stocks, the volatility of material supply and sales determines whether the results will be representative for production. In cases where fluctuations over the year are limited, either indicator can be used though "value of production" is preferred in cases where fluctuations are substantial.

As production and sales values are in current prices, a comparison over time requires a deflation of these values to prices of the reference period. If it is compiled in the country, the Producer Price Index (PPI) is the most appropriate deflator for the production and sales values recorded in the IIP. It should be noted, however, that the PPI gives less accurate results for industries with homogeneous products or major inputs deflation than when volume or quantity changes is used. The reason for this is that producer prices are differentiated between buyers, and companies treat these realized producer prices as extremely confidential. In general, producers provide list prices rather than the actual transaction prices for the PPI.

2.2.1.2 Measurement of inputs

When reliable measures of industrial output are not available, the measurement of inputs can sometimes be used to approximate industrial production. As mentioned earlier, labour input as well as material inputs can be used as proxies for production.

Labour input is an indirect measure of output and can assume various forms. The most accurate measure is the number of hours worked. The number of full-time equivalent jobs is the second best measure of labour inputs and the number of persons engaged is the third best option. Ideally, the labour input data needs to be adjusted whenever the relationship between labour input and value added changes, but such information is normally not available. Labour input most accurately reflects output in large-scale enterprises with many unskilled or semi-skilled workers. In smaller businesses with highly skilled labour, it is an unsuitable indicator, as employers want to retain the skills and knowledge in their enterprises and do not rapidly reduce the labour force. They will also have difficulty finding the right workers once the business environment improves again. The country's labour laws and regulations also influence the quality of this indicator. In many countries it is difficult to lay off workers. This results in slow adjustments of the labour force and in such cases, shifts in labour force size is not a good indicator of production changes.

In some industries, a single or just a few important materials – in terms of their share in material input costs - are used to manufacture a wide range of articles. This applies, for instance, to printing, where paper is converted into books, brochures and many other products of various sizes. The quantity of paper used is a good indicator for the output as it is the major input. Similar situations exist in the manufacture of furniture and fixtures where wood and board products are major inputs, for various cement products where cement gives a good indication of production volumes and in the manufacture of clay bricks and tiles of various shapes and sizes where the amount of clay used is a good indicator for measuring production in a common unit of measurement.

It is not possible to recommend a single method that is best for measuring industrial production for the IIP. For industries producing homogeneous products, the quantity of output is the most appropriate indicator. For those with a wide range of products based on a few critical inputs, measuring those inputs is the best option while determining the value of output or sales may be the better solution in other cases. However, when no price suitable deflators are available for those activities where

values should ideally be used to measure output or sales, the use of labour inputs (preferably, the number of hours worked) is probably the only feasible option.

2.2.1.3 Methods of obtaining industrial production in volumes

There are two methods to obtain industrial production estimates in constant prices (volumes). One method is deflating values of production or sales and the other is the direct measurement of volumes such as quantities of output, material inputs or a labour force indicator.

The resulting production volumes for a given period are compared to the production volumes of the base period, using the same indicators for each enterprise. These ratios of change form the elementary indices in the IIP system. Aggregate average indices can be calculated using weights sets that describe the relative importance of all enterprises within an activity and, subsequently, of all activities within the total.

2.2.2 Weight set preparation and baseline information

Two different weighting approaches are available for aggregating the elementary indices of the IIP, namely weighting by commodity or by industry. Industry weighting is more accurate than commodity weighting because it focuses on the way industrial production is organized and the data is accessible.

Commodities are produced by companies that can be classified under different activities, depending on whether the selected commodity is a primary or a secondary product. As the definition of share of production is relative rather than absolute, it may be that a substantial part of production is classified under various activities where it is not, in fact, a principal product. Unless specific measures such as centralized product trading corporations or excise duty systems are in place, which in most countries only apply to a few products, comprehensive data on the production of selected commodities can only be obtained through more extensive surveys. It is much more difficult to estimate the total output of all major commodities than of major industries, and it is therefore also more difficult to select the correct commodities for an index system. It is furthermore often difficult for enterprises to provide information on value added at the product level, which is required for a proper weighting system.

The selection of suitable units for industry-weighted index systems is much more straightforward. Such systems use weights at different levels of aggregation, based on

the assumption that the commodities, establishments and industries selected in the index calculation are representative.

The selection of the respondents for a sample survey often relies on the 70/70/70 rule, which is as follows:

1. Select the largest activities (for example, at 4-digit ISIC), which together account for at least 70 percent of industrial production in terms of value added. For the IIP, value added should be measured at basic prices;
2. Within each of the selected activities, select the largest establishments which together account for not less than 70 percent of the value added of an activity;
3. For each of the selected establishments, select product lines that account for 70 percent of the total production of the establishment. As value added is not calculated at this level, base it on the output value, but again at basic prices.

This procedure generally results in the selection of 150 to 200 establishments, with a total of some 250 to 300 production lines. In most activities, two or three enterprises account for the bulk of production and those will therefore be selected. There are some activities, however, with many similar-sized operators, for example, in garments, textile and furniture manufacturing. The top ten companies may not even reach the 70 percent target specified in rule 2 above. In such cases, it normally suffices to use the five largest units to measure developments over time. There is no need to enumerate more than ten units in a stratum[5] at any time.

The most common source for the data used to calculate the weights for the IIP derives from industrial surveys or censuses carried out by the statistical office, such as annual surveys of manufacturing industries. Alternatively, administrative data on value added tax (VAT) or the sales tax system can provide the necessary information for sample selection and the preparation of the weights.

If an annual industrial survey is carried out or if VAT data are used, it is possible to generate annual weights sets for the IIP. The introduction of annual weights sets improves the accuracy of the index series, as the structural changes that take place in the country's industrial activities are automatically taken into account in each new weight set. The use of a chained index allows measuring the annual production

5 Depending on the size and distribution of a country's industrial activities, indices could be prepared at the country level or at a regional level. In the latter case, the 70/70/70 rule should be applied to the results at the regional level, which are subsequently aggregated to the total at country level.

changes, taking the most recent changes in the country's industrial structure itno account.

One of the problems with the long-term use of base-weighted index systems is that they generate a downward bias in the results. The longer the same base period is used, the larger the bias that is introduced. It is therefore important to ensure regular updates of the weights for the IIP/PPI system. The compilation of annual weights sets would not only solve the problem of a bias, but would also allow for improved estimates of current developments using chained indices or a combination of Laspeyres and Paasche indices for the IIP and PPI, respectively, so that the Fisher value indices can be calculated.

As mentioned earlier, there are two possible sources for updating the weights sets annually, namely (annual) surveys of industrial production and the value added tax system. Both sources have their advantages and disadvantages, which are discussed below.

Industrial surveys are usually conducted by the statistical office, and the surveys' full data are therefore available to the statistical office for any given purpose. The main problem with the surveys is their timeliness, as there is a lag of up to two years between the end of the reference year and the availability of the results. In addition, these surveys are not conducted annually in many countries. Furthermore, the list frame for the survey may not always be up-to-date as the maintenance of business registers is cumbersome and not always carried out systematically. On the other hand, information on the value added and gross output is available for all companies participating in the survey, and gross output for individual product lines is also generally reported in these surveys.

The main advantage of using VAT data is that they are made available in a much more timely manner, often on a quarterly basis. Because they are collected under the tax law rather than the statistical law, the statistical office may face problems in obtaining the data. To avoid such problems, aggregate data rather than company details could be requested. The minimum data needed are value added and gross output recorded in the VAT system at the 4-digit ISIC, either on an annual or a quarterly basis. Furthermore, the list of the largest new companies is needed to verify whether all large companies are included in the index system for the upcoming year as well. The main requirement for a proper comparison of the aggregates compiled in

the VAT and the equivalents in the index systems is that all companies are classified in the same ISIC group in both systems.

Given that the timely availability of new base weights is very important, the use of VAT data for annual weights sets is recommended. As tax authorities usually implement strict procedures to enforce compliance with the law, the number of companies omitted from their system is likely to be lower than from an independent business register maintained by the statistical office. This will allow for the timely inclusion of new enterprises in the index system.

2.2.3 Index type

Selecting the index type to be used depends on both theoretical and practical conditions. A Paasche type index benefits from its up-to-date weighting structure, but generally cannot be produced in a timely manner. Compared to a Paasche index, the Laspeyres index can be produced in a timely and cost-effective way. However, as the industrial structure is fixed in the base year of the series, the index becomes less accurate over time as the relative importance of the different industries changes. This generally results in a downward bias of the index, which increases with time.

As previously mentioned, the compilation of a chained index series whereby each year's data is evaluated against the weights of the previous year, and the annual indices chained, results in a series with weights representing the current industrial structure. This series has the advantage of timeliness and overcomes the bias problem inherent in a regular Laspeyres index.

To reiterate, the IIP is a short-term indicator and should only be used as an estimate of production until comprehensive data from a recent industrial survey can be incorporated into the GDP estimates.

2.2.4 Data requirements

The data requirements of an index system are twofold, namely recurrent information on production (and prices) to be collected as an ongoing activity, preferably at monthly or quarterly intervals, and the information needed to construct the weights sets for the system for each year for which such data are available.

Under the present methodology, the production volume for a given company in each period is compared with the production in the base period and the ratio between the

current production and the base year values, the so called *production ratios*, form the basis for the index calculations.

To establish an accurate system, the companies' detailed survey information (for the selected units only) in terms of *gross output* and *value added* should be used. Value added shares are used for the production index weights.

The selection of activities, companies and product lines to be included in the index follows a simple procedure in principle. Only the most important activities that together account for more than 70 percent of gross output and value added are selected. As the other activities are of lesser significance, together accounting for not more than 30 percent of total output or value added, their omission from the index system cannot distort the results in the short term (up to five years in practice), as such a distortion would only occur if the average growth rates of the selected activities varied significantly from those not selected, which is highly unlikely.

Within the selected activities, the largest companies which together account for more than 70 percent of output and value added are again selected. In industries which consist of many units of similar size (and characteristics), such as the garment and furniture industries in many countries, it is not necessary to meet the 70 percent criterion as long as a sufficient number of respondents (for example, five or ten) is selected.

In many instances, companies have a single product line but in others, multiple product lines are common. If a company has several product lines, the most important ones in terms of output are considered, with a maximum of three product lines per company. Those three will normally add up to more than 70 percent of the total value of the company's output. A product line is not the same as a product. A soft drink manufacturing company often only has a single product line where all drinks are produced. They consist of a mix of sugar, water and some concentrate only, often with carbon dioxide added at bottling time. For production purposes, the output is the volume of all soft drinks combined and not the production volumes of individual flavours of these sugary drinks. If the company also were involved in the bottling of water, this would constitute a second product line as the input-output structure of this output differs completely from that of soft drinks.

2.2.5 Data collection and data processing

The IIP is generally compiled as a monthly series, but some countries produce it on a quarterly basis. The principal reason for recommending a monthly data collection and publication cycle is that the indices are very important for the Quarterly National Accounts (QNA). The publication of the QNA is time bound and if a country aims to participate in the IMF's Special Data Dissemination Standards (SDDS), the quarterly data collection for the initial estimate of the QNA would generally not be completed in time. Using a monthly data collection schedule, the initial estimates can be based on the actual figures for the first two months of the quarter and on estimates for the last month[6].

As a monthly data collection schedule puts additional pressure on the respondents due to the increased frequency of visits, the changeover to the new system could also be used to update data collection methods for those companies ready to embrace them. Various options are available such as the electronic submission of the data in .pdf (or .xls) format as attachments to electronic mail or, if the electronic infrastructure of the statistical office allows, direct data entry by respondents into the IIP/PPI database hosted on the organization's website.

The use of regular mail or fax to send hard copies of the information could be an intermediate step for those companies not yet comfortable with using electronic data transmission. The first year of the new system could be used to train individual companies in the use of the data transmission option most suitable for their specific situation. It must be noted that the IIP system only covers the largest companies and that many already extensively rely on the internet for certain parts of their business. Exploiting those competences for IIP data collection makes it easier for respondents to change over and it simultaneously reduces the work load for the statistical office. Direct data entry into the database would eliminate all intermediate activity altogether, while the use of fixed formats for data transmission (in either .pdf or .xls) allows for the mechanical transformation of the data to the database. In both cases, most of the editing of the data can be performed by the respondents at the time of preparation as various edit checks can be built into the forms.

---

6 When using the PRIMA system for data management of the indices, the forecasting facility provided in its analytical module can be used for this purpose.

2.2.6 Publication targets

The production indices by major activity can either be prepared for the entire country or into broad regions for some type of sub-division. In case of regional indices, the overall size of the sample needs to increase in order to maintain the required accuracy.

If the production and/or producer price index series are expected to provide separate estimates on production for local use and exports, it must also be verified that the industrial establishments selected for the index systems have an adequate representation of both types of production.

The detail used during the compilation process is not directly related to the publication format. To ensure the accuracy of the calculations, it is best to use the four-digit ISIC codes as the lowest weighting level. However, the limited number of units in some of those activities or in some regions may preclude publication at this level of detail and only the 2-digit level data may be released. This does not differ from common practice in CPI computations whereby information is processed at the individual COICOP item level (hundreds of products), while the results are only given for the 11 COICOP functions. At the same time, detailed information is available whenever required for use in the preparation of the national accounts and other economic statistics.

2.2.7 Producer price indices and the IIP

Occasional references to the Producer Price Index (PPI) have been made in the previous chapters. In many countries, the preparation of the IIP and PPI are organized as two separate activities, often in different departments of the statistical office. In most cases, this is a historical development, with the industrial statistics department in charge of the IIP and the prices department is responsible for the PPI. There are several strong arguments in favour of combining the IIP and PPI's data collection and processing into a single operation. These are enumerated below.

First, the respondents to both surveys will be the same companies when a single national index series is prepared, or the PPI becomes a sub-set of the IIP sample

when regional production indices[7] are prepared. Companies that have the largest production volumes are also those able to set or influence the prices.

Second, the amount of recurrent data collected from the companies is minimal once the product lines for the IIP and the pricing items for the PPI have been identified, but the frequency will be high, preferably monthly. In many countries, data collection for such surveys is conducted through field visits. As only senior company staff is generally allowed to provide information to visitors, they need to participate in these visits, which mostly are not by appointment. That is, the senior staff must frequently interrupt their work (24 times a year for monthly surveys) which can result in response fatigue. Combining the surveys reduces the response burden by half and, consequently, also has a positive effect on the response rate. The company visits should be used to 'wean' the companies off interviews and encourage them to use electronic platforms to provide the necessary data, thus reducing the burden for the companies. An enquiry based on response by email can also be designed in such a way that data can be automatically copied to the IIP/PPI database, thereby reducing data entry and transcription errors in the statistical office.

In the descriptions of the weights sets for the IIP, the first and second stage weights are based on the value added of the activities and companies, while the product lines within a company are only weighted based on gross output, because value added information is not available at this level of detail. Value added is used because it is the appropriate measure for GDP from the production side. Gross output is the appropriate measure for the measurement of prices, as prices represent the total value of goods, not just the value added in the industrial production process.

As mentioned, the companies selected for the PPI are the same as those used for the IIP, and the general information on the monthly or quarterly questionnaires for both is therefore common. Only production and price information differ for the two surveys. For the IIP, the production volume by product line continues to be used; for the PPI, the prices of the most common transactions for some very specific products within those product lines are collected.

---

7 The PPI is only prepared at national level.

3. The experience of Tanzania

3.1 Background

The United Republic of Tanzania is a developing country with an estimated population of 42.8 million in mid-2011. It is located in East Africa, south of the Equator, and borders Kenya and Uganda to the north, Rwanda, Burundi and the Democratic Republic of the Congo to the west, and Zambia, Malawi and Mozambique to the south. The economy has been growing at 7 percent on average since 2000, heavily relying on agriculture, tourism, mining and manufacturing in recent years. The manufacturing sector comprises mining and quarrying, and electricity, gas and water supply comprises 13 percent of GDP in 2007. Since the manufacturing sector generates more jobs and revenue, the Tanzania Development Vision (TDV) 2025 and the National Strategy for Growth and Reduction of Poverty emphasize industrialization as the primary objective in the country's future.

Under the One UN programme, UNIDO's intervention in the United Republic of Tanzania covers a number of integrated activities with various components, including industrial statistics. One of the activities is the improvement of the industrial statistical system so coherent, relevant, timely and reliable industrial statistics can be provided for knowledge-based industrial policymaking. The objective of the Industrial Statistics sub-component is to enhance the capacity of the Ministry of Industry, Trade and Marketing (MITM), the National Bureau of Statistics (NBS) and the Confederation of Tanzania Industries (CTI) to collaborate in the collection, compilation, analysis and dissemination of industrial statistics. The expected outcome of the Industrial Statistics sub-component is an enhanced capacity of MITM, NBS and CTI in collecting, maintaining and disseminating industrial statistics and its analysis for policy-relevant issues. In 2008, the overall framework of project implementation was agreed. Between 2008 and 2010, the Annual Survey of Industrial Production 2008 (ASIP 2008) was conducted, processed and published[8], and the related database was set up.

8 United Republic of Tanzania, *Annual Survey of Industrial Production and Performance 2008 (Tanzania Mainland), Statistical Report*, Dar es Salaam, September 2010, and United Republic of Tanzania, *Annual Survey of Industrial Production and Performance 2008 (Tanzania Mainland), Analytical Report*, Dar es Salaam, November 2010.

In September 2008, the project was launched with an assessment mission by UNIDO staff and international experts to better understand the needs of different database users in Tanzania and to elaborate a plan for implementation of the proposed activities by different parties. Following the positive response from the ministry, the second mission was conducted in January and February 2009 to follow up on the progress and to organize a training course in industrial statistics, emphasizing the latest version of "International Recommendations for Industrial Statistics" released in 2008. International industrial statisticians also provided trainings for 33 users and 16 supervisors to jointly conduct the annual industrial survey involving MITM, NBS and CTI. An international consultant was in charge of the third mission in June-July 2009 to follow up on the progress and to discuss the proposed questionnaire and other methodological documents and working plans for conducting industrial surveys with different parties in Tanzania. It was agreed that the joint industry survey would cover the year 2008, and it was entitled 'Annual Survey of Industrial Production 2008', or ASIP 2008 for short.

ASIP 2008 started, as planned, in October 2009, with the data collection beginning on 5 October 2009. The survey included establishments with 10 or more employees from four sections of B, C, D and E of ISIC Rev.4. The data collection period lasted for over two months and was completed at the end of 2009. In the following three months, data cleaning and data editing were conducted. The fourth mission headed by industrial statisticians took place between 25 April and 8 May 2011 to check the quality of the data derived from the survey, to generate tables from the survey data and to prepare the statistical report of ASIP 2008. Two reports were generated from ASIP 2008, a statistical report by an international industrial statistician, and a report by a national consultant, Dr. Semboja, together with staff members from the Ministry of Industry, Trade and Marketing (MITM), Tanzania. The main findings from the survey were well documented in these two reports. More importantly, ASIP 2008 set up a database for more substantive uses, such as the construction of a new index of industrial production.

3.2 Production Index in Tanzania

3.2.1 Introduction

The IIP in mainland Tanzania[9] - the Production Index of Manufacturing Industries (PIMI) - uses 1985 as a base while the PPI uses 1995 as a base. For the PIMI, the volume base year was also shifted to 1995 when the PPI was first introduced. This shift in the PIMI's volume base does not represent a rebasing of the series because the weights were not changed, and therefore continues to reflect the relative importance of the different activities in 1985.

At present, data collection for PIMI and PPI is carried out on a quarterly basis, although monthly production and price data are collected. The questionnaire currently being used for data collection contains all the necessary information for both the PIMI and PPI indices in the new data processing system (PRIMA) that has been installed. The main changes required to the questionnaire relate to the change-over from quarterly to monthly data collection, which simplifies the questionnaire in general. Furthermore, some of the data repeated every month in the questionnaire can be asked in a more straightforward way, keeping reporting requirements of repetitive information to a minimum.

For short-term indicators to be accurate, it is important for the reference period to be as recent as possible. The present PIMI and PPI series have base periods that are over 15 years old and it is obvious from the comparison of the indices' base weights information in the 2008 ASIP that the industrial structure in the country has changed considerably over this period of time.

The ASIP 2008 provides a good source of information for the rebasing of both index series and the construction of an integrated system of production and producer price indices for the economy's industrial sector. Currently, the PIMI and PPI cover manufacturing activities only, but using the 2008 ASIP information, the indices can be extended to include mining, electricity and water supply. In principle, the same system can also be used for various services for which monthly volume indicators are available (e.g. hotel bed nights, telephone call minutes, etc.), as these activities are important for the Tanzanian economy and no such data are presently being collected

9 In Zanzibar, no production index is being compiled, but data on the production volumes of important industrial commodities is collected and published.

on them. Rather than creating a separate infrastructure for such activities, their inclusion in the currently available series is a much more efficient approach.

3.2.2 Datasets for ASIP 2008 and the existing PIMI and PPI series

3.2.2.1 Annual Survey of Industrial Production - 2008 (ASIP)

The 2008 ASIP provides the new base weights for both the PIMI and PPI. Therefore, the information contained in this survey needs to be streamlined with both these index series. ASIP covers mining, electricity and water supply (ISIC Rev.4 codes B, D and E) as well as the manufacturing industries (ISIC Rev.4 code C), which are presently covered in the PIMI and PPI.

ASIP uses a simple coding structure based on the region ID (region, district, and locality) and a serial number. The serial number is not the same from one survey to the next. ASIP 2008 uses ISIC Rev. 4 to classify activities.

3.2.2.2 Production Index of Manufacturing Industries (PIMI)

The PIMI (commonly referred to in other countries as the Index of Industrial Production – IIP) contains information on the quarterly production of selected commodities by select companies in Tanzania. The system in its current form dates back to 1995. It is maintained as an Access table, but operated as a spreadsheet[10]. Part of the data preparation, namely the conversion of various production details to the PIMI standard, is done outside the system in a temporary spreadsheet. The PIMI database contains quarterly data, although the actual information is obtained on a monthly basis. The aggregation of monthly to quarterly data is also done outside the database as part of the conversion process.

The PIMI dataset does not have unique identifiers (codes) for companies or products. It uses an alphabetic code for location (region) and ISIC revision 2 for activities. The companies included are generally those that have been prevalent since the mid-1990s, although some have changed and the old or new name has been added to others in brackets in the name field. During the 1990s, many state-owned companies

[10] In a database system, the number of fields (columns) in the table is fixed and new information is added as new rows. In a spreadsheet, new information may be added anywhere, but this is mostly done by adding columns or sheets in the system. For PIMI, a number of columns is added to the table every quarter.

were privatized and they often changed names in the process. For some, three or four names can be found spanning the period from 1985 (the PPI reference base) to now.

In general, the PIMI dataset contains a column for each quarter indicating whether the information provided in the table has been received from the field or was carried forward from previous quarters as estimates (or was estimated, but not carried forward as the same data). For 2008, this information could not be found for the second quarter, prompting the "in general" qualifier mentioned above. Knowledge about the nature of the data in the database is important so a company or product line can be properly classified as closed or in operation. When the same non-zero production, employment and turnover data are repeated, the information is obviously carried forward, but in many cases, only production data is indicated. As these are often fairly small numbers, there is a good chance production levels do not differ noticeably from one quarter to the next when looking at this factor only. If the turnover, which is quoted in thousands of shillings, is the same as well, chances are that the data is repeated, since identical outcomes from one period to the next are very unlikely for large values.

The PIMI dataset includes the (monthly) quantity and value of production in 1995, the base year. As the index is a direct average of the available information and is not re-weighted for relative coverage in the different industries at group level, this is all the information required to build the 1995 weights set. The data derived for PIMI cover the years 2007 to 2011. The 2007 data are quarterly while the later data are monthly. For consistency purposes, the 2007 data have been converted into monthly data by equally distributing the quarterly data over three months.

3.2.2.3 Producer Price Index (PPI)

The PPI contains data on a monthly basis for a sub-set of the companies covered in the PIMI, but with a dozen or so establishments that are not found in the PIMI. The PPI has been collected since 1995, and although prices are collected on a quarterly basis, the three monthly quotations of each quarter are processed separately. Thus, a monthly time series of prices is available. The data for the three years 2007 – 2010 cover 296 products reported by 135 establishments. This list only includes companies that were operating during this period, unlike the PIMI list which includes all establishments on the original list.

The dataset does not include codes for the companies or products, but the standard region codes form part of the information received. ISIC Rev. 2 is used to code the activities and products. The data received cover the years 2007 to 2009. The coverage of the PIMI and PPI in terms of establishments is somewhat different. The full list of the products included in PIMI and PPI, along with the number of quotations, is given in the table below. It should be noted that the descriptions have been generalized. Specifics for each product group are given in their "quality" definitions as explained later. The main omissions from the PPI are export commodities such as fish products, black tea and cured coffee. In the PIMI, some consumer goods categories are missing (e.g. bread, butter). The major differences are highlighted in *Table* 2.

Table 2 Comparison of coverage by PIMI and PPI by ISIC Rev.4 item code

Item code	Commodity	Number in PIMI	Number in PPI	Item code	Commodity	Number in PIMI	Number in PPI	Item code	Commodity	Number in PIMI	Number in PPI
0149A	Bee wax	1		1394C	Nylon Twines		1	2393A	Ceramics	1	
0149B	Honey	4	1	1394D	Poly Ropes		1	2394A	Cement	3	3
1020A	Fish products	15		1399A	Canvas	3		2394B	Lime	2	1
1030A	Canned fruits and vegetables	12	5	1399B	Canvas goods	1		2410A	Aluminium circles/sheets	1	1
1030B	Fruit and vegetable juices		3	1399C	Sisal Carpets	1		2410B	Corrugated Iron sheets	3	1
1040A	Vegetable oils and fats	42	14	1399D	Textile bags	2	1	2410C	Galvanized Pipes	2	1
1050A	Milk products	11	4	1430A	Knitted garments	3		2410D	Rolled steel	13	
1050B	Butter and ghee		3	1512A	Leather goods	2		2410E	Steel sheets/billets	3	1
1061A	Maize flour	14	6	1520A	Beach sandals	6		2410F	Mild Steel Round		4
1061B	Rice	3		1520B	Shoes	4	5	2511A	Louvres	1	
1061C	Wheat flour	14	7	1610A	Timber	19	10	2511B	Steel structures	2	
1071A	Biscuits	10	2	1621A	Plywood and building boards	3	2	2593A	Agricultural tools	3	5
1071B	Bread		8	1701A	Paper	2	2	2599A	Aluminium wares	2	2
1072A	Sugar	5	4	1702A	Paper products	6	2	2599B	Crown corks	2	
1072B	Sugar confectionery	1	1	1709A	Paper tissues	1	1	2599C	Metal containers	2	
1072C	Molasses		1	1920A	Petroleum fuels	8		2599D	Wire products	5	8
1079A	Black tea	26		2011A	Industrial & Medical gases	1	4	2599E	Washbasin		1
1079B	Blended tea	5	6	2021A	Pyrethrum extract	1		2599F	Weld Mesh Sheet		1
1079C	Coffee beans, cured	15		2021B	Insecticides & pesticides - liquid	2		2599G	Drums		1
1079D	Instant Coffee	5	3	2021C	Insecticides & pesticides - powder	3		2599H	Expanded metal sheet		1
1080A	Animal feeds	4	9	2021D	Mosquito coils	2	1	2640A	Radios & R/Cassettes	1	
1101A	Spirits	3	1	2022A	Paints	10	4	2710A	Electrical motors	2	
1103A	Bottled beer	5	21	2022B	Thinner		1	2710B	Transformers	2	4
1103B	Chibuku	1	1	2023A	Soap and laundry / toilet detergents	20	12	2720A	Auto batteries	1	6
1103C	Malt	1	1	2023B	Dental Cream		1	2720B	Dry cells	1	1
1104A	Soft drinks	17	8	2029A	Adhesives & industr. Detergents	1		2720C	Electric Bulbs		1
1104B	Drinking Water		3	2029B	Safety matches	2	1	2732A	Electrical cables/wires	2	4
1200A	Cigarettes	1	6	2100A	Pharmaceutical products - capsules	2	3	2818A	Elect cookers	1	
1200B	Tobacco, cured	3		2100B	Pharmaceutical products - syrups	5	5	2818B	Industrial machines	3	5
1311A	Cotton yarn	9	1	2100C	Pharmaceutical products - tablets	6	5	2818C	Machine tools & equipment	1	
1313A	Woven fabrics	16	12	2211A	Auto tubes	1	3	2920A	Motor bodies and trailers	8	
1391A	Knitted fabrics	6		2211B	Auto tyres	1	9	2930A	Radiators	1	
1391B	Mosquito netting	1		2220A	Plastic articles	17	12	3091A	Bicycles	1	1
1392A	Blankets	2	2	2220B	Polythene bags	4	3	3100A	Foam mattresses	12	17
1394A	Fishnet & products	3	1	2310A	Glassware	1	1	3290A	Brooms/brushes	1	
1394B	Sisal ropes and twines	5	1	2392A	Concrete Tiles		1				

A total of about 100 different commodities or product lines are listed in the table above. The number of commodities in PIMI and PPI is both roughly 70. It should be noted that not all items listed in Table 2 are actually produced at present, as several companies on the PIMI list have closed down since the system was designed. For compatibility, the entire list has been used to design the codes so older data can be added to the database at a later stage, if desired.

3.2.3 Construction of weights

In order to set up the IIP system in Tanzania using the ASIP 2008 database, three missions by international consultants were organized. During the first mission in April 2011, the situation in Tanzania was assessed, including the index situation, the data requirement, the hardware preparation and training for counterparts. In preparation for this first mission, an outline of the proposed IIP system was elaborated and submitted to UNIDO in mid-February. The ASIP 2008 report was used as a source of data for the regional analysis.

Upon arrival in the field, information on the current activities of NBS with regard to industrial production and producer price indices was collected. This included a review of the systems used as well as the grouping of regional data into broader zones.

Weights sets – new base period. ASIP 2008 represents the main source of the data required for the weights sets. The necessary data for the calculation of the indices is given in aggregated form, as presented in *Table 3* and *Table 4* (originally from Tables 27 and 28 of the statistical report of ASIP 2008, respectively). The information in *Table 3* by company (for the selected companies) and sub-aggregates at the 4-digit ISIC level (for all companies in the ISIC group) provides information on gross output and value added required for the weights sets of the price index and production index, respectively.

Table 3 Production account by industrial activity (ISIC Rev.4) 2008 *(*millions Tanzanian shillings)

ISIC R4	Description	Persons engaged (persons)	Gross output	Intermediate consumption	Value Added (basic prices)	Compensation of employees	Taxes on productionless subsides	Gross operation surplus
B	Mining and quarrying	2,618	357,977	201,005	156,972	27,600	484	128,888
C	Manufacturing	107,388	4,711,376	2,913,993	1,797,383	342,413	107,301	1,347,670
10	Manufacture of food products	43,792	1,109,701	630,456	479,245	66,001	6,452	406,792
Of which								
102	Processing and preserving fish and similar products	3,878	274,670	202,461	72,209	8,405	1,076	62,728
104	Manufacture of vegetable and animal oils and fats	1,107	73,828	41,097	32,730	1,782	56	30,892
106	Manufacture of grain mill products, etc.	2,205	104,473	69,963	34,510	1,744	63	32,703
107	Manufacture of other food products	36,344	647,069	312,338	334,731	53,588	5,239	275,904
11,12	Manufacture of beverages ad tobacco products	12,593	1,104,511	700,075	404,436	71,425	96,502	236,509
13,14,15	Manu. F textiles, wearing apparel and leather products	13,430	206,594	119,623	86,971	14,242	948	71,781
16,17	Manu.of wood, paper and similar products except furniture	2,641	33,226	18,789	14,437	4,908	214	9,315
18	Printing and reproduction of recorded media	3,695	160,808	89,725	71,083	17,198	256	53,630
20,21	Man. Of chemicals including pharmaceutical product	4,687	427,145	242,530	184,616	18,584	275	165,757
22,23	Manufacture of rubber, plastics and non-mineral products	6,685	773,230	518,583	254,647	89,361	4,074	161,213
24,25	Manu. of basic metals and fabricated metal products	3,213	222,239	173,396	48,842	7,790	-2,394	43,446
27,28,29,30	Manu. of electric, transport and other machinery and equip.	925	67,129	52,330	14,799	3,308	20	11,471
31	Manu. of furniture	2,613	25,722	17,189	8,533	2,775	21	5,737
32,33	Other manufacturing	13,114	581,070	351,296	229,774	46,820	935	182,019
D	Electricity, gas, steam and air conditioning supply	6,375	756,901	602,845	154,056	61,165	-24,075	116,966
E	Water collection and treatment	1,241	33,608	17,555	16,053	4,566	-6,699	18,186
Total large industry		117,622	5,859,862	3,735,398	2,124,464	435,744	77,011	1,611,709

The data in Table 4 (Table 28 in the ASIP 2008 statistical report) at company level, provide basic information for the product lines operated by the various companies as well as the relative importance of domestic and export sales if it is decided that indices of that type need to be prepared.

The 2008 ASIP includes 729 establishments in the mining, manufacturing, electricity and water supply industries. A quick analysis shows that about 300 establishments with over 50 employees each represent 92 percent of total output and the largest 192

units (with 100 or more employees) generate 83 percent of total output. The latter share is large enough to ensure reliable results at the national level in mainland Tanzania.

Table 4 Sales of product by category and destination, 2008 (million Tanzania shillings)

CPC V2	Description	Exports	Local	Total
1	Product of agriculture, horticulture and market gardening	6,068.2	123,550.6	129,618.8
2	Live animals and animal products (excluding meat)	0.0	7,415.2	7,415.2
3	Forestry and logging products	0.0	863.0	863.0
4	Fish and other fishing products	20,112.1	0.0	20,112.1
14	Metal ores	191,189.9	363,175.7	554,365.6
15	Stone, sand and clay	139.8	10,197.2	10,337.0
16	Other minerals	2,838.7	8,790.4	11,629.1
17	Electricity, town gas, steam and hot water	0.0	93,155.9	93,155.9
18	Natural water	3,398.2	33,063.9	36,462.1
21	Meat,fish, fruit, vegetables, oils and fats	137,984.2	421733.9	559,718.1
22	Dairy products and egg products	583.5	7,188.5	7,772.0
23	Grain mill products, straches and starch products; other fodo products	309,218.0	148,207.9	457,425.9
24	Beverages	2,398.9	410,373.0	412,771.9
25	Tobacoo products	0.0	165,636.7	165,636.7
26	Yarn and thread; woven and tufted textile fabrics	9,855.4	105,166.1	115,021.5
27	Textile articles other than apparel	29,704.3	104,920.3	134,624.6
28	Knitted or crocheted fabrics; wearing apperal	1,350.8	19,379.5	20,730.3
29	Leather and leather products; footwear	914.5	410,790.8	411,705.3
31	Products of wood, cork, straw and plaiting materials	1,267.1	9,777.7	11,044.8
32	Pulp, paper and paper products; printed matter and related articles	1,987.4	141,518.9	143,506.3
33	Coke oven products; refined petroleum products; nuclear fuel	1026.2	23,926.4	24,952.6
34	Basic chemicals	0.0	198,646.9	198,646.9
35	Other chemical prodcts; man-made fibres	20,477.7	129,858.7	150,336.4
36	Rubber and plastics products	109.5	19,283.7	19,393.2
37	Glass and glass products and other non-metallic product n.e.c	0.0	811,872.4	811,872.4
38	Furniture; other transportable goods n.e.c	287.9	133,536.7	133,824.6
39	Wastes or scraps	0.0	35.0	35.0
99	Not indicated	66,723.6	1,029,659.8	1,096,383.4
	Total	807,635.9	4,931,724.8	5,739,360.7

The distribution of employment in 2008 by region is provided in

Table 5 (original Table 5 of the ASIP 2008 statistical report). These regions have been grouped into the five zones used by the NBS. Information on employment size for each region and the zones, along with the grouping of the regions into five zones, is presented in Table 6.

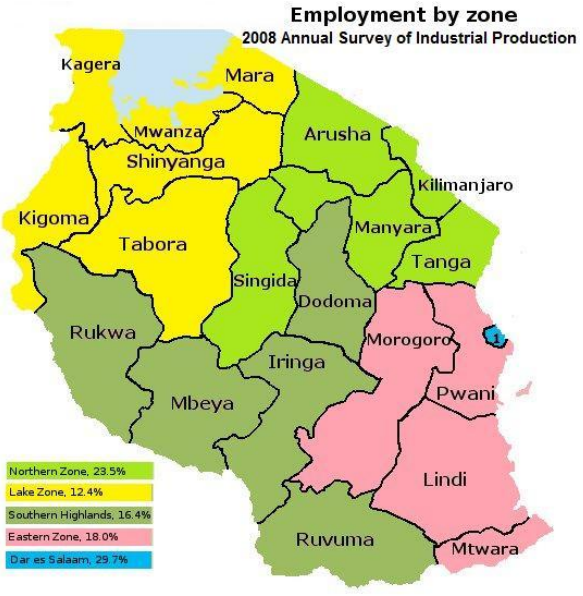
Table 5 Number of establishments by region and employment size, 2008

Region	Employment size					total
	19-Oct	20-49	50-99	100-499	>=500	
Dodoma	4	1	1	2	-	8
Arusha	10	18	7	10	6	51
Kilimanjaro	7	5	4	5	3	24
Tanga	4	4	6	10	3	27
Morogoro	6	5	1	4	6	22
Coast	3	2	-	3	-	8
Dar es Salaam	167	80	51	74	11	383
Lindi	3	1	2	-	-	6
Mtwara	1	3	1	1	-	6
Ruvuma	1	3	2	1	-	7
Iringa	4	3	3	5	4	19
Mbeya	4	6	1	5	3	19
Singida	17	6	3	1	-	27
Tabora	1	3	1	2	-	7
Rukwa	-	2	3	-	-	5
Kigoma	3	1	2	2	-	8
Shinyanya	-	4	3	3	1	11
Kagera	7	8	2	7	2	26
Mwanza	10	7	10	7	5	39
Mara	4	1	3	5	-	13
Manyara	7	5	-	-	1	13
Total	263	168	106	147	45	729

Table 6 Employment grouped into regions in Tanzania, 2008

Number of persons engaged by geographical region, June 2008

Zone	Cd	Region	Number	Percentage	Zone with share	
1	2	Arusha	13,900	12.5%	Northern	
1	3	Kilimanjaro	5,541	5.0%	Northern	
1	21	Manyara	985	0.9%	Northern	
1	13	Singida	583	0.5%	Northern	
1	4	Tanga	5,057	4.6%	Northern	23.5%
2	18	Kagera	5,227	4.7%	Lake	
2	16	Kigoma	512	0.5%	Lake	
2	20	Mara	1,401	1.3%	Lake	
2	19	Mwanza	4,558	4.1%	Lake	
2	17	Shinyanga	1,379	1.2%	Lake	
2	14	Tabora	704	0.6%	Lake	12.4%
3	1	Dodoma	419	0.4%	Southern Highlands	
3	11	Iringa	13,891	12.5%	Southern Highlands	
3	12	Mbeya	3,024	2.7%	Southern Highlands	
3	15	Rukwa	274	0.2%	Southern Highlands	
3	10	Ruvuma	391	0.4%	Southern Highlands	16.2%
4	6	Coast	615	0.6%	Eastern	
4	8	Lindi	132	0.1%	Eastern	
4	5	Morogoro	18,994	17.1%	Eastern	
4	9	Mtwara	258	0.2%	Eastern	18.0%
5	7	Dar es Salaam	32,988	29.8%	Dar es Salaam	29.8%
	22	**Total**	**110,833**	**100.0%**		100.0%



In this grouping, the regional distribution of employment is fairly even, with one-eighth of total employment in the Lakes Zone at one extreme, and 30 percent in Dar es Salaam at the other. The other three zones lie between those two extremes. To arrive at useful results for PIMI at the regional level, the units selected for the indices need to reflect this regional coverage of activities.

If the production and/or producer price index series are expected to provide estimates of production for local use and exports separately, it needs to also be verified that the industrial establishments selected for the index systems have an adequate representation of both types of production.

In order to check whether the coverage of zones and industry is adequate, the shares of all ASIP 2008 respondents has been expressed as shares in the total within the five zones as well as their percentage in the two-digit ISIC groups.

During the mission, a substantial amount of information on the existing PIMI and PPI systems was obtained and analysed. Furthermore, additional data from the ASIP 2008 was also obtained. With this information, an initial weights set for the new PIMI (regional) and PPI systems was developed in a workshop environment with the staff of NBS and other partners. The full weights set could not be finalized during this mission as some of the information was not readily available, but was completed in a subsequent mission.

In addition, the ASIP 2008 data was analysed to identify those establishments significant enough to be included in the new index system. Several considerations played a role, including ensuring that the regions were properly represented and that all large units were selected.

The survey sample includes all those units whose share in the zones' output is more than 0.25 percent (188 units). In addition, other units with a high share in production at the 2-digit ISIC level (31 with more than 5 percent) have been included. This leaves 510 units in ASIP, of which some have reported a large employment size, but very low gross output (and thus very low gross output per worker). This was the result obtained using the 70/70/70 rule. However, eight establishments were not covered in the selection above which reported more than 500 workers and equivalent labour costs. Their value added was very low, probably due to reporting or editing errors. These units were also added to the selection. This then resulted in an overall selection of 227 units for the MIPI and PPI systems. These establishments represent 84 percent of employment and 94 percent and 93 percent of gross output and value added, respectively. In other words, those units provide a suitable (actually, fairly high) coverage of activities in the industrial sector, which ensures that there is no need to include more units in the PIMI/PPI system.

3.2.4 Review of data quality of the ASIP and PIMI/PPI datasets

The ASIP data were analysed in various ways to check for major errors that might exclude some eligible establishments from the selection. In addition to the summary information by establishment, gross output and value added per worker and the ratios of these indicators to their averages were reviewed as well. All ratios that were more than ten times the average or less than one-tenth of the averages were flagged. The ratio between the highest and the lowest gross output per worker was more than 300,000, which is far too high and indicates problems in the establishment data on both extremes. Only the question whether any larger units in terms of employment size have been excluded due to editing problems in the ASIP 2008 dataset is of importance for the construction of the sample set for PIMI.

After the first mission was completed, the establishments included in PIMI and PPI were matched to the furthest extent possible. Matching had to be done by name, and the names used in the two datasets were not always the same. This mapping identified 122 companies in both index series. For 14 companies in the PPI, no matches were found in the PIMI series, while there were 283 entries in the PIMI for which no PPI equivalents were found. It should be noted that there are 405 establishments in the (1985) PIMI dataset of which 27 were marked “closed”, and it seems that other companies in the PIMI dataset had also closed before the PPI system was finalized after 1995.

The 2008 data for the quarterly PIMI surveys suggest that there are more companies that have closed since the series was started, as no turnover data for 2008 was available for another 200 establishments. For some companies, information was clearly copied from previous quarters because no responses were received. Furthermore, some establishments reported large variations in turnover from quarter to quarter. Others reported turnover for only a few quarters while there were units that reported no turnover but did report either production quantities or employment size - or both. The discrepancies have been checked against the original questionnaires and the errors that were found have been corrected in the datasets.

3.2.5 Data preparation: coding and classification

A general outline of the data preparation steps was provided at the early stage of the project, when the actual data had not been reviewed. The overall activity can be divided into several steps, some of which are necessarily sequential, that is, one step

has to be completed before another can commence. This requirement applies in particular for the data conversion to the PRIMA software system, which can only be carried out after all coding and mapping has been concluded. Under the best of circumstances, the preparation and conversion of data requires several weeks of work, and as outlined below, not all conditions of the Tanzanian dataset were necessarily the best.

3.2.5.1 Linking the three surveys

The records in the three surveys had to be matched. While the datasets were not very large (together only about 1500 records), there was a number of specific challenges. Since no other information was available, the final matching had to be done by name. The first step in the matching process was to assign a code to all enterprises based on location and activity. This code was created as follows:

a. Zone code (1, digit, values 1 to 5)

b. Region code (2 digits, values 01 to 21)

c. ISIC code (4 digits, ISIC Rev. 4, 4 digits, i.e. PIMI and PPI data were first recoded to ISIC Rev.4).

The total length of this part of the ID is 7 digits. At a later stage, a 2-digit serial number was added to complete the company code.

3.2.5.2 Geographical codes

As mentioned, the first part of the ID is a geographical code. The codes are those used at the NBS. They are provided in the table below, along with the number of companies in each region in the three datasets.

Table 7 Regional distribution of companies in ASIP 2008, PIMI and consolidated database

Zone	RegCd	Region	Name	Zone Name	ASIP08 count	PIMI count	Company Count
1	2	Aru	Arusha	Northern	51	38	64
1	3	Kil	Kilimanjaro	Northern	24	17	29
1	4	Tan	Tanga	Northern	27	38	44
1	13	Sing	Singida	Northern	27	2	27
1	21	Many	Manyara	Northern	13	1	14
2	14	Tab	Tabora	Lake	7	3	7
2	16	Kig	Kigoma	Lake	8		8
2	17	Shi	Shinyanga	Lake	11	12	12
2	18	Kag	Kagera	Lake	26	16	29
2	19	Mwa	Mwanza	Lake	39	30	45
2	20	Mar	Mara	Lake	13	11	16
3	1	Dod	Dodoma	Southern Highlands	8		8
3	10	Ruv	Ruvuma	Southern Highlands	7	3	7
3	11	Iri	Iringa	Southern Highlands	19	17	22
3	12	Mbe	Mbeya	Southern Highlands	19	17	25
3	15	Ruk	Rukwa	Southern Highlands	5	1	5
4	5	Mor	Morogoro	Eastern	22	16	24
4	6	Coa	Coast	Eastern	8		8
4	8	Lin	Lindi	Eastern	6	1	6
4	9	Mtw	Mtwara	Eastern	6		6
5	7	Dar	Dar es Salaam	Dar es Salaam	383	182	410
99	**99**	**ZZZ**	**Total**	**Mainland Total**	**729**	**405**	**816**

The combination Zone-Region was chosen as the geographical code to simplify processing. This code will always be three digits without leading zeroes. If region alone is used, the code will be of mixed length (one or two digits) or with a leading zero for the low numbers, which is problematic in terms of processing. Using the zone code as the geographic code might result in difficulties for the regional allocation of units in the analysis.

3.2.5.3 ISIC – activity coding

Conversion of the ISIC codes from revision 2 to revision 4 is a three-step procedure, considering that the cross-classification tables available from the UN Statistical Division are from revision 2 to revision 3, then from revision 3 to revision 3.1 and finally from revision 3.1 to revision 4. It is not an entirely automatic process, as the changes in activity classification also involved the merging of some activities in the old classification to create a new activity in some cases, and in other instances, involved dividing existing activities into (parts of) multiple new activities. Hence, there were three possibilities at each conversion step:

a. Simply recode an old activity into a new one (automatic exercise)

b. Merge two or more activities into a single new one (automatic exercise) and

c. Subdivide an activity into two or more new ones (manual exercise).

There are 81 groups in manufacturing activity in ISIC Rev.2, resulting in 307 combinations in the cross-classification because of multiple entries. Out of those 307, only 14 had a unique relationship to a new code in revision 3, while 13 more groups were merged into new groups. For 280 of the codes in revision 2 (90 percent of the total, which are not all present in the Tanzanian activity lists), the new code had to be assigned manually.

- The conversion from ISIC Rev.3 to ISIC Rev.3.1 only involved 8 manufacturing activities that were distributed over new activities. While it had to all be done manually, the low number of activities involved simplified it.

- The conversion from ISIC Rev. 3.1 to ISIC Rev. 4 referred to 127 manufacturing activities in ISIC Rev 3.1. A total of 344 combinations were available of which 47 were automatic conversions and the remaining 85 percent had to be done manually.

As the PIMI and PPI datasets were largely the same and both used ISIC Rev.2, a temporary step was to link these two datasets first using ISIC Rev.2 and the next step was the creation of the equivalent ISIC Rev.4 codes for all units in either of the two datasets and the new company code was subsequently generated.

Using the first seven digits of the company code describing location and activity, the key parts of the extended PIMI dataset (there were 14 PPI entries that did not have a counterpart in PIMI and were thus added, hence extending PIMI) were extracted and this set was added to the ASIP dataset. Sorting by code and name resulted in a table with quite a large number of companies being identified and cross-referenced. As might be expected, this was an entirely manual process as automatic name matching only works if the names are identical in spelling.

Company names in the three datasets generally referred to those used during the time of the datasets' establishment (in 1985, 1995 and 2008). During this period, the Tanzanian economy experienced major structural changes from an economy largely operated through parastatal companies in the eighties, to a mixed economy in the nineties to a currently largely private economy. Thus, many companies changed hands (often, several times) which generally resulted in name changes. The PIMI and

PPI datasets used secondary (older) names in brackets for a number of companies, but many companies could nonetheless not be matched on the basis of their names. The PIMI records added to the ASIP dataset also included the 2008 production and employment data as recorded in PIMI.

3.2.5.4 Product codes

At present, the PIMI and PPI use a single field to describe pricing product and product line. In addition, there is a field for unit of measurement. This information is too detailed for proper aggregation, on the one hand, and not detailed enough, on the other, for all processes in a fully automated system. This latter requirement was further emphasized when trying to insert all necessary information in the additional detailed format that was designed to replace the manual calculations of total quantities now being entered outside the system in spreadsheets. Therefore, the different levels of detail for the product and quality descriptions as applied to the Tanzanian data in the new database system are described below.

Following the PRIMA naming conventions, there are items (products or commodities) which are defined at a broad level, and for each item, there are a number of qualities. Qualities are described by various characteristics and can have a number of levels of detail. For the CPI, for instance, a good, systematic description is the Structured Product Description (SPD) as developed for the 2005 round of the ICP project. This defines each pricing item in meticulous detail and, when applied properly, allows for adjustments of newly introduced qualities (packaging, size, etc.) to be carried out automatically.

For the PIMI and PPI, the following system has been implemented:

a. Item (product, commodity): a broad grouping of outputs, such as 2100A Pharmaceutical products, 1040A Vegetable oils and fats, 2599A Aluminium hollowware, 0149B Honey, etc. Item codes consist of the ISIC code plus a letter, making it an alphanumeric code which can accommodate leading zeroes. While this coding scheme limits the number of items per activity to 26, it is highly unlikely that more than ten items are defined in any activity group.

b. Quality: A number of fields used to describe the item in adequate detail. In general, there will be a sub-division of the description if more than one characteristic is mentioned in the same field. The actual composition depends

on the item and varies between activities. Common methods of describing qualities include:

1. Product line
2. Product size
3. Material used/produced and
4. Packaging, etc.

In addition to these quality descriptions, the unit of measurement and the conversion factor to the standard unit of measurement for the item for each establishment. It should be noted that aggregate units of measurement do not need to be the same for equivalent items in different establishments, as this largely depends on the way the given company measures its output. For instance, in the U.S., textile manufacturers measure output in feet or yards, while European manufacturers measure it in metres. Depending on the company's parent organization, the standard unit of aggregation may differ. As production for a given company is always compared over time and expressed as ratios, the unit of measurement for an individual company does not play a role in any of the calculations. The re-defined list of items was presented in Table 2. This is not a final list, as some of the items listed can probably be further grouped into similar groups. A few examples of quality descriptions are:

- 2100A Pharmaceutical products - tablets
  a. Type of product: paracetamol, aspirin, ibuprofen, etc.
  b. Concentration of active ingredient: paracetamol 250, paracetamol 500, etc.
  c. Packaging size: box of 100 strips of ten, bottle of 1000 tablets, etc.
- 2100B Pharmaceutical products - liquids
  a. Type of product: paracetamol, aspirin, ibuprofen, etc.
  b. Concentration of active ingredient: paracetamol 250, paracetamol 500, etc.
  c. Packaging size: 12 bottles of 50 ml, 6 bottles of 100 ml, etc.
- 2100C Pharmaceutical products - capsules
  a. Type of product: paracetamol, aspirin, ibuprofen, etc.

b. Concentration of active ingredient: paracetamol 250, paracetamol 500, etc.

c. Packaging size: box of 100 strips of six, bottle of 100 capsules, etc.

- 3100A Foam mattresses

  a. Density of the foam (high, medium or low)

  b. Finishing (foam only or finished mattresses)

  c. Size: l * b * h (= volume)

- 1081A Maize flour

  a. Milling grade (wholemeal, fine/white meal)

  b. Packaging size.

For each item/product, there is a unique set of characteristics to describe it at two levels of detail. The first level, which is broad, is the product line; the second level is the detail of the pricing item. In the three examples given above, separate product lines are only identified in the first example in terms of the form of delivery of the medicine (tablet, liquid, capsule, etc.). These are prepared by different machines/ production processes and the items have to therefore be differentiated in the production index. The other items do not have such a description, unless the secondary step for foam products (putting a cover over the foam or not) is considered as such. In general, the easiest is to treat this type of product finishing as a quality difference rather than classify it as a separate item. Along the lines outlined above, all existing information has been converted into a set of items / quality descriptions that define the production item (product line) at the item level and the pricing item at the final level of detail.

The quality code is simply a serial number for each company, without an effort to assign the same serial number for identical items across different companies, as it would be complicated to maintain this. Furthermore, there are many product groups where quality as defined here is unique and available only at that specific company. That is the case, for instance, for all branded products, as brand is always part of the quality description and specific to a single producer or corporation. Thus, each product in the economy can be identified as the quality of an item produced by a company or establishment, or a set of three codes: company-item-quality.

3.2.5.5 Quality codes for PIMI products

Applying the scheme outlined above to the PIMI dataset for the first three months of 2008 was not without problems. The first step was the assignment of company codes. Followed by ISIC codes, which were converted into Item codes at a later stage, the ISIC code of the main activity is embedded in the company codes, and it was separated because it was to be one of the codes used for assigning the item codes.

The next step was to review the initial codes assigned as item codes against the commodity descriptions provided in the file. At the same time, the quality codes (serial numbers) were also assigned for all items reported by a company. Based on the production information, the multipliers used to obtain output data in uniform units of measurement were also calculated. In Table 4, several of the companies listed in Table 3 are repeated, but with the additional information. It should be noted that the information for the first unit is converted from kilos into tonnes, while that for the second one is converted from bottles (with different contents) to cubic metres (or thousand litres). The data for a textile mill were changed into thousands, while the output of an equipment supplier remained in single units. The volume of soap was converted from boxes (of 7 and 14 kg, respectively) to metric tonnes while the final entry may or may not be correct.

3.2.6 ASIP 2008 coverage

During the process of assigning company codes to this file, it was determined that 12 of the companies listed in the detailed production file neither appeared in the main PIMI file nor in the ASIP 2008 dataset. These 12 companies were added to the list.

While the omission of eight manufacturing companies will not have much of an influence on the overall totals for ASIP, the missing gold mines have a major effect. According to various statistical sources, Tanzanian gold production was about 39.8 metric tonnes in 2008. The only gold mine included in the ASIP is Resolute's Golden Pride mine. In 2008, this mine accounted for just over 11 percent of gold produced in the country. In other words, nearly 90 percent of gold production was not covered, although some of it was included in the PIMI reports reviewed, despite not being transferred to the database or spreadsheets.

This omission has a major in influence on the share of mining in the country's total industrial production. Assuming that the other gold is produced at the same prices as Resolute's implicit price, which was about 12 percent below the average sport price of

gold in 2008, and excluding other income generated from copper and silver at one of the other mines, the industry shares are estimated as shown in Table 8, as reported in ASIP 2008 and after the abovementioned adjustments.

Table 8 Industry shares in 2008: before and after gold mine adjustment (%)

Industry	Before	After
Mining and quarrying	4.2	21.9
Manufacturing	82.7	67.4
Electricity	12.3	10.1
Water and sewerage	0.7	0.6
Overall change		22.7

As can be seen, using these conservative estimates, total industrial production is nearly one quarter higher than reported in ASIP, with a major shift in the relative importance of sub-components. Mining's share rose by about 16 percentage points, while the other activities decreased in shares but not, of course, in actual output.

As mentioned earlier, it seems that a number of establishments were identified in the ASIP 2008 that may have been duplicated. If we assume that most of the establishments are actually duplicates, the gross output of manufacturing is over-estimated by between 100 and 140 billion shilling. Although this is only about 1.5 percent to 2.5 percent of gross output overall, it is about double that in terms of employment. For some of the specific activities, the difference will be substantially higher though. To establish a proper basis for the PIMI weights for 2008, these duplicates have been removed.

The various adjustments to the 2008 baseline discussed in the previous sections are necessary to establish a proper weights set for the index series. For instance, if the basic information for the missing gold mining companies is not included, it is not possible to assign an appropriate weight to these units. Similarly, any company not included in the base period does not have a weight assigned in the index system. Thus, although production and price relatives will be calculated from period to period, the resulting ratios are multiplied by zero (the weight) if these commodities were not included in the baseline.

A summary of the adjustments of manufacturing establishments to the 2008 baseline is presented in Table 3 and Table4. Also, the PIMI included a substantial number of establishments that were not operating or for which no production data were available for the entire year. In the table below, only those PIMI establishments are

listed for which either a match was found in ASIP 2008 or for which production data were available for at least one quarter in 2008. All other establishments listed in PIMI were not consolidated into the main 2008 dataset. Following all the matching exercises and further adjustments, 759 establishments were identified that were operating in 2008. Of these, only 667 were listed in the ASIP 2008 dataset. For the 101 units for which (partial) data were available in the PIMI, gross output and employment were estimated for the full year and added to the consolidated database. For the extra units, only physical production quantities for the first quarter of 2008 were available, so the gross output could not be estimated as of yet.

Table 9 Number of manufacturing establishment by survey

Survey		PIMI			
		Yes	No	Extra	Total
ASIP 2008	Yes	190	477	-	667
	No	101	-	8	109
	Duplicates	-	-17	-	-17
	Total	291	460	-	759

As indicated earlier, a weights set derived on the same basis as that for 1995, namely directly from the respondents, has also been prepared for 2008. The limitations of such a weights set have also been indicated. As some of the required adjustments to the 2008 base data are quite significant, this new weights set will not properly measure the changes in the country's industrial production. All companies, items and qualities have been coded, even though not all information required is always available. In other words, the coding structures are complete, but not all assigned codes are correct yet. The data in the detailed item specification file is not always adequate to assign conversion factors. In the database, a one (or one/thousandth) has been entered as a conversion factor where the actual one could not be derived. At the same time, the aggregate converted data for all items of all companies has been entered so that it is possible to re-create the 2007 and 2008 PIMI indices using the original PIMI weights set for 1995 and the revised limited one for 2008. It should be noted that the 1985 weights used for PPI and the average prices for that year are not available in any of the datasets provided, so for now, the PPI can only be compiled with the limited 2008 base year weights derived from / for PIMI.

To the extent the available data permitted, all conversions have been completed. The details of the summary tables presented here have been entered into a single excel workbook for reference by NBS. In addition, the weights, identifiers (companies, items and qualities), production data (2007 and 2008) and price data (2007 to 2009)

have all been converted into the new PRIMA database format. As already mentioned, the multipliers assigned for individual products have been the weakest, as the proper multipliers could not be reconstructed in a significant number of cases. In those cases, one or one-thousandth has been used as a first approximation. However, this did not lead to useful data for testing and this set of data has therefore not yet been included in the sql backup that will be sent to NBS for installation on their system.

4. Summary and Recommendations

As may be clear from the description in the previous chapter, the preparations of a consistent database for the PIMI and PPI in Tanzania and the corresponding weights sets took much effort and time. When first establishing an IIP/PPI system, the conversion of old index information is not necessary, but the preparation of the weights sets requires the same effort. The actual work required for the preparation of the weights depends on the specifics in each country. The description in the previous chapters of the work needed in Tanzania gives a good indication of the types of actions required in the preparation of a weights database.

The same applies for countries where historical data need to be converted. The main argument for investing in converting existing information into the new database system is that there is no break in the series and that the entire data series remains available for analysis. This is especially important for regressions and projections, which require a minimum length of the series, generally five years or so, before they can be applied. When starting a new series, it will take that long before all analytical capabilities can be used. With compatible historical data on the indexing database, these facilities can be used immediately.

For the preparation of the weights set, the following information is needed:

1. The full dataset for the industrial survey database to be used for the weights sets. This should include as a minimum the establishment name, location, activity, employment size, wage bill, gross output (by product line), intermediate consumption and value added, plus the grossing up and raising factors used for each individual respondent;
2. The classifications and coding lists used for activity, location and products; and
3. Listing used as the survey frame as well as membership lists of industry associations and chambers of commerce and industry, the large taxpayers register (VAT), etc.

Before the actual preparation of the weights set can be started, the following editing steps need to be carried out:

1. Calculate wages, gross output and value added per worker for each establishment as well as gross output and value added ratios. Sorting the

dataset from large to small on each of these characteristics successively, analyse the large and small outliers and introduce corrections if the error can be located. It is very unlikely that the gross output or wages per worker differ by more than a factor of ten between the highest and lowest ratios, and outliers are errors caused by problems in enumeration or data entry that need to be corrected. Similarly, negative value added is unlikely, unless in price-controlled activities. In any case, negative value added for a company needs to be adjusted as it implies that the larger the output of the company is, the lower its contribution to GDP, which is illogical;

2. Verify the coverage of the survey against the various membership lists to ensure that all the large industrial enterprises are covered. Special attention has to thereby be given to the more recent additions, as the lists used as the survey frame are often not up-to-date. As can be observed in the case of Tanzania described above, most gold mines were not covered in the survey as these had only been established in the last few years before the 2008 survey and the survey list frame was not up-to-date;

3. Ensure that the units of measurement are consistent over time, from the base period onwards for each commodity within each establishment. If inconsistencies are found, the reporting unit used in the current data should be used throughout and older information with different units of measurement have to be converted into the new ones using the standard conversion units or, where that is not possible (e.g. cubic metre of timber into tonnes), the most accurate approximation; and

4. Define the main parameters of the planned index series in terms of coverage, publication detail of activities and geographic detail. As the publication of information on individual respondents is not permissible under most statistical acts and regulations, it is necessary to check whether the planned publication detail fulfils this condition. Furthermore, any distribution, such as the regional one, should be fairly even. For instance, if activities are to be grouped into two regions for publication purposes and the resulting weights are 90 percent for one region and 10 percent for the other, the regional distribution does not add any useful information as it is far too skewed. In the case of Tanzania, as described in section 3.2.4, the regional shares for the five regions are between 12.5 percent and 30 percent, whereby the average is 20 percent. This provides a very good basis for analysis. However, the regional

distribution of individual activities is not as even and the publication detail for activities cannot be as detailed at the regional level as it is at the national level. To prevent the disclosure of confidential information, the activity breakdown of manufacturing at regional level is at the 2-digit ISIC detail, while it is at the 4-digit ISIC detail at national level.

When back data also needs to be converted into the new database system, the following additional information is needed:

1. Monthly or quarterly data by establishment and commodity for all enterprises and periods for which the information needs to be converted;
2. Classifications or descriptions used in the dataset for location, activity and product and unit of measurement for all establishments covered;
3. Prepare cross-classifications, if required, between the codes or descriptions used in the existing IIP dataset and the new weights system and dataset. Subsequently, re-code all old information to the new classifications and code sets. This is important because modifications to various classifications, such as ISIC, are carried out every 15 to 20 years and the information in older data series needs to therefore be re-coded before the information can be merged into the new database.

The existing data also needs to be edited before being merged into the new database. This may sound odd, as the information was used earlier to generate the indices. It should be noted, however, that previous use does not always imply correct information. For instance, in most cases the old system would use the change in average or total production of a given good during a specific period rather than the current recommendation of the change in production by company. In the present circumstances, a decimal error in reporting by an enterprise will be noted immediately as the individual data are analysed. However, in the old system it was often the change in total or average production that was compared from time to time. It would not always be very obvious that these averages contained errors, as any error is diluted by the other correct responses.

A second form of editing would be required for information in the old dataset which has not been used so far. As it might be used in the new database, either for the calculation of the indices or for analytical purposes, this data also needs to be evaluated. For instance, employment numbers, wages or turnover may have been

collected, but because the information was not used, no editing was carried out. Calculating and comparing changes in these numbers for each enterprise over time may display outliers. Similarly, the computation of wages and turnover per worker and their comparison over time as well as between enterprises can help identify errors that need to be corrected or replaced by imputed values to ensure consistency in the data uploaded to the new database.

The outline for the creation of weights sets and the conversion of existing data into a new database system holds for any database system to be implemented. In Tanzania, the PRIMA software with its MySQL database[11] has been used to implement the new IIP/PPI system. As it is one of only few integrated indexing packages available and is provided in the open source environment, a brief description of its features and capabilities has been included in the Annex.

[11] PRIMA uses the open source MySQL database system as a standard, but the software can also be used with other database systems. It has in the past been installed using an Oracle database, for instance.

Bibliography

International Monetary Fund, *Producer Price Index Manual*, Washington, 2004

PRIMA System, *Installation manual*, Tilburg, 2002-2012, limited distribution for software users, adapted for each individual application

PRIMA System, *Desktop user manual*, Tilburg, 2002-2012, limited distribution for software users, adapted for each individual application

PRIMA System, *PDA user manual*, Tilburg, 2002-2012, limited distribution for software users, adapted for each individual application

PRIMA System, *Indexer manual*, Tilburg, 2002-2012, limited distribution for software users, adapted for each individual application

PRIMA System, *Analytical module manual*, Tilburg, 2002-2012, limited distribution for software users, adapted for each individual application

United Nations, *International Recommendations for the Index of Industrial Production 2010*, Statistical Papers, Series F, No. 107, New York, February 2010 (draft)

United Republic of Tanzania, National Bureau of Statistics, *Annual Survey of Industrial Production and Performance, 2008 - Statistical Report*, Dar es Salaam, September 2010

United Republic of Tanzania, National Bureau of Statistics, *Annual Survey of Industrial Production and Performance, 2008 - Analytical Report*, Dar es Salaam, November 2010

United Republic of Tanzania, National Bureau of Statistics, *Quarterly Production Of Industrial Commodities: 2004 – 2009 – Mainland Tanzania*, Dar es Salaam, February 2011

Appendix: The PRIMA software

The compilation of indices is fairly complicated and is better done in database systems than in spreadsheets. An open source index system is available for the development of the new production and producer price index series. This PRIMA (PRice Index MAnagement) system has been implemented in a number of countries for the compilation of different price and production indices (e.g. CPI, PPI, IIP, WPI, unit-values of external trade, etc.)[12]. If desired, PRIMA allows for direct data entry by respondents on the website/in the database; it can be used with hand-held devices or smart phones for data capture and verification and comes with an extensive analytical capability, including the possibility to generate forecasts and projections.

As part of the installation, extensive documentation is available and training in all aspects of its use is provided. Continuous support is provided by email or remote access, but from the experience with long-term users so far, support requirements are minimal. In the past nine years of use in Ghana, for instance, only general software upgrades have been provided and the statistical service has been able to publish its CPI on schedule (the second Wednesday of the month at 10 AM) without interruption every month.

One of the main advantages of the PRIMA software suite over alternatives is that it allows an unlimited number of weights sets for any index series, so that it is easy to include annual weights for the IIP and PPI. Furthermore, it also enables the calculation of various types of indices (Laspeyres, chained Laspeyres and Paasche) in parallel, giving flexibility to the analysts who can select the most appropriate option for a particular purpose.

A.1 A brief introduction

PRIMA is the abbreviation of **PR**ice **I**ndex **MA**nagement system, an open-source indexing system. As indicated above, it has been applied in several countries to produce different types of indices for prices, production and unit values.

a. Modular design

The system consists of three parts/modules:

1. Data entry module: data entry and editing
2. Index modules: index calculations
3. Analytical module: tabulation, analysis and publication

b. Generic structure

The system is data-independent. It takes the structure for a particular application from the structure of its (hierarchical) weights file.

---

[12] PRIMA has been in use in Maldives and Ghana since 2002 and is being installed in various countries in the Ecowas region for their harmonized CPI. It is also used in Dubai, Palestine, Sudan and pilot applications have been installed in a dozen other countries.

• It allows for an unlimited number of price quotations

• It allows for an unlimited number of pricing items, qualities, stores, regions, etc.

• It allows for weekly, monthly or less frequent price observations

• It has eight levels of aggregation from the Item-level upwards

• It allows for separate weights files for each period and purpose

• It generates Laspeyres, Paasche or chained indices as required

• It accommodates up to 12 different weights sets within one weights file so that indices for different groups (low income, high income, salaried workers, etc) can be calculated separately. Alternatively, 12 monthly weights can be provided so that the seasonal availability of products can be considered in the weights.

The system is flexible and designed for various kinds of use. For example, data entry can be flexible, using off-line PDAs (Personal Digital Assistants) or using a desktop system within a local network, or through the Internet using PDF forms.

Index calculation can, in principle, be initiated from any point on, while the results can be tabulated by either producers or users through the Internet. The actual implementation of a particular index system depends on the available infrastructure and facilities within a country and the need or desire to use the Internet for data entry and/or dissemination. Various features can be set by the system administrator. User interfaces are standardized and have easy-to-use graphics (windows/web). They can be modified by the system administrator for use in a particular situation.

Data entry and editing procedures are application-specific, that is, they need to be developed and adjusted for each individual system. The index calculations and tabulation and publication functions are generic, that is, they are the same for all applications.

The most important part of the system is the set of weights which has to be developed in a strictly hierarchical way. While this weights system is part of the indexer, much of its information is copied to the data entry system as the same data hierarchy can be used to present and print the data entry forms.

A.2 Approach used in the PRIMA system

A.2.1 Data entry module

The main features of the data entry module are:

• The data entry screen is an exact representation of the items and their qualities priced within each store or in a market in a given period (week, month, quarter).

• The prices for the previous period are displayed along with the product/ quality information for ease of reference.

• The supervisor can set an acceptable price range. The default[13] range is 25 percent around the previous price for the same item in the same store. All prices outside that range will be marked. The data entry operator can only proceed after explicitly accepting the new price, which is labelled in the database as out-of-range. The supervisor can then review the list of all transactions for the selected period that are out of range, review and edit them. If data entry is done in the field using hand-held devices, the reason for any major change as given by the seller is recorded and the information is displayed along with the suspect data in the review screen.

• When a change in quality or source/store occurs for an item, it can be introduced in the database at the time of data entry. In a separate pop-up screen, the descriptive information on the new quality or source can be provided as well as current and previous prices of this new quality. The ratio between the price of the old and new quality in the previous period is then used to impute prices for the new quality in all previous periods, so that there is a price for any current quality in any previous period that might be selected as a new base period, e.g. in a chained index. The underlying assumption for the procedure described is that the price movement of a new quality is the same as that of the quality it replaces for the earlier periods when prices for the new commodity were not available. This is the standard assumption also used in the CPI manual.

• The new qualities will be used in later questionnaires and data entry schedules, so that the information available is always up-to-date. For CPI systems, the International Comparison Project is currently developing a standardized set of item descriptions. If this information is available in the appropriate format, it can be incorporated in the PRIMA system to describe the qualities of the pricing items in full.

• The data entry module is directly linked to the new questionnaire. The latter can be generated for printing - in the form of a PDF file - or in electronic format for copying to a hand-held device. The questionnaire for the market or town includes only those qualities, stores and items for which prices have to be collected during the selected week. This is based on the distribution of outlets and the frequency of price collection for individual products over the monthly or quarterly cycle as specified in the system layout for the application.

• After the data entry and editing process has been completed, the data is ready for index calculations. The final steps in the editing process to be carried out by the supervisor include the review of discrepant data and of non-response, and if this

[13] The purpose of the range check is to catch data entry errors, not to vet price changes. This is the reason that PRIMA does not have separate ranges for different product categories. Any large price change is also captured in this list of discrepant prices and the reasons given for such significant changes can be used to understand and analyse the development of inflation more accurately and faster.

option has been activated, the imputation of prices for common items[14] across the markets and regions.

A.2.2 Indexing module

The index calculations are fully automatic. After starting the indexer, a selection screen appears. Here, the appropriate selections have to be made, and the indices subsequently calculated. The following items need to be selected:

• Survey set: The indexer is generic. Different index datasets may be present (CPI, PPI, WPI, IIP, etc). If only one survey set is loaded, it is automatically selected.

• Weights set: Multiple weights sets may be present for a system. For instance, a PPI may have both weights based on output and value added. The CPI may have weights sets using the traditional commodity grouping, another one using COICOP and a further one giving a grouping designed for analytical uses, for instance, by local and imported items. If only one weights set is available for the selected survey set, this is automatically selected.

• Index type: A choice between Laspeyres, chained and Paasche types is available. Paasche indices have to be calculated using weights for the current period or the latest available period. If only a single base period is available for a survey set, the Paasche index degenerates to Laspeyres by default. The same applies to the chained index as it will look for the weights set before, but nearest to the year for which the data are processed, which in case of a single dataset is the original base period only.

• Archive: This option, when selected, will write all indices calculated for the period, from the individual item in a store up to the overall total, in a set of archive files. This is the dataset that is used in the analytical module of the system. In case of large datasets, archiving takes some time. Archiving is included as an option so that the initial rounds of index calculations are not delayed by the waiting times. Only the last version of the indices is kept in the archives, anyway, so no information is lost when carrying out this procedure only once on each month's dataset.

• Seasonal weights: Some countries calculate monthly weights for seasonal products in the CPI and sometimes in WPI or PPI. When this option is selected, the 12 [Weights Columns] available in the system are allocated to the 12 months. Data for January are automatically weighted with Column 1, those for February with Column 2, and so on. If the option is selected while not all 12 columns contain weights, the indices for the corresponding months logically show up as zero.

• Weights Column: By default, this is 1; it only needs to be set if more than one weight is available for calculations (e.g. low, medium and high income groups). After the selections have been made, the base period and the periods for which the indices

[14] Common items in the CPI are those items for which a single pan-territorial price applies. This includes postage stamps, mobile phone prepaid cards, various government fees, but in some countries also the price of fuels, newspapers, etc. This feature reduces data entry errors and ensures that prices are available in each location. The common items are only priced in the selected market or town, and thus only show up on those questionnaires.

need to be calculated have to be selected. In case indices for more than one group need to be generated, the index calculations need to be repeated for each of these groups. If the option [Seasonal Weights] has been selected, this option is not available and blocked.

• Base period: First, the base period needs to be selected from the list of available options. The base period is defined in terms of the pricing reference. For instance, *1997-09* as a base period means that the prices of September 1997 are used as reference in the calculations. Similarly, *2004-00* means that the average prices for the year 2004 are used as reference. These average prices for the calendar year are calculated on-the-fly and need therefore not to be present in the database. There is also an option to define any base period for the index, for instance, the fiscal year July 2003-June 2004, or the accounting year April 2004-March 2005, or a specific shorter period such as August-November 2004. In addition to providing weights for the selected period, it is also necessary to calculate and upload the average prices for this selected base period. In both the weights set and the database, the period should be indicated as *year-13* or, in this case, *2004-13*.

• Periods: All periods with data available for the weights set selected are shown in this data block. One or more periods can be selected for index calculations. The actual number of periods that can be selected as a maximum depends on the size of the weights set and the maximum memory size allocated to the indexer. In general, between 20 and 30 periods can be run simultaneously, but for very large systems, this may be reduced to five only. The results are stored in an archive. New data will be added and re-calculated data will replace those already in the archive file.

• Calculate indices: After all selections have been completed, this option should be used and the calculation of the indices is carried out. Progress of the calculations is shown at the top of the screen. If the [Archive] option has not been selected, index calculations generally do not take more than one or two minutes, depending on the size of the index system and the number of periods selected. If the [Archive] option is selected, the operation for multiple periods may take up to 15 minutes.

• View / Export: The results of the calculations can be viewed by selecting this option. If chosen without calculations, it will show the latest data in the memory. Data can be viewed at different levels of aggregation, from full detail at Item-level for each individual item or establishment (depending on the index selected) to overall total only. The export option transfers the data at the level of detail shown in the view to a HTML or Tab-separated text file (.txt) for import into and use in any database, word processor or spreadsheet application.

A.2.3. Analytical module

The analytical module allows the preparation of reports, based on the archive of the indexer, in a wide variety of formats and layouts. A number of derived characteristics can also be calculated.

The analytical module is Internet-based. When the indices, but not the underlying prices, are uploaded[15] regularly onto the specified website, users from all over the world can access the information and obtain their own specific index series.

All indices can be displayed up to the lowest level, but the actual maximum level of detail selected for the particular session is determined in the first steps of the process. If greater detail is required, the selection process needs to be repeated from the start.

All information can be filtered at all levels to generate specialized, detailed index series for specific purposes. Filtering can be done by region(s) or for different product groups, or for selected items only (fresh produce, imported goods, goods or services, etc.) in any combination. The options described below are available for unfiltered as well as filtered data, unless conditions for regressions determine otherwise.

Any or all periods for which data are available in the index set can be selected. For de-seasonalization and regressions, consecutive periods need to be selected. For both activities a minimum number of periods is also required. This is three years for de-seasonalization and 40 periods for regressions.

When filtering data, the selected information can be aggregated anew, or the aggregates for higher level indices can be copied from the data file. In the former case, only the weights for the selected items are included.

Three display formats are available: html table format, in comma-delimited text format for easy import in spreadsheets and other applications, or as jpg graphs.

Features are added upon demand. The following have already been implemented:

1. Month-to-month inflation;
2. Year-to-year inflation;
3. Average year-to-year inflation;
4. De-seasonalized indices (using X11);
5. Inflation forecasting model (using TRAMO); and
6. contribution to inflation for the period by individual items and groups.

The forecasting model can project forward for one to 12 periods from the last selected period. It uses up to five external variables as regressors, of which at least one needs to be present. In each system, common variables such as exchange rates, crude oil prices and their indices are available in the database. These can be selected using a pull-down menu. Other regressors can be inserted as required from simple text files. In these cases, it is necessary to ensure that the first period in the file matches the first period in the regression, as no further checking takes place. When a selected

---

[15] If the system has been set in such a way that the MySQL database containing the index information is on the organization's server, the [Archive] option in the indexer module automatically writes the most recent information to the proper location for use by the analytical module.

regressor is not available for the entire period of the regression, the operation will be terminated with an error message.

A.3. PRIMA concepts

A.3.1 Item-level

The Item-level is the pivotal level in the system. It is always defined implicitly in the weights file of an application as the lowest level for which weights are provided.

The number of levels below the Item-level is in principle unlimited. At the lowest level, the price quotations are converted into price relatives (relative to the earliest base period of the system). The structure of a dataset from price quotation to Item-level may, for instance, look as follows:

For each item in the system, prices are collected

- for a number of qualities;
- in a number of stores;
- in a number of markets/towns/regions.

At each level up to the Item-level, all price relatives within a "group" are averaged, using a simple geometric or arithmetic average, as specified in the set-up information for the application.

A.3.2 Quality

The quality is normally the lowest level in the system. Price quotations are collected for particular qualities in each location.

Quality is defined here in a very practical manner, so that field workers do not have to use their own judgements in most cases, thereby reducing errors: "*If two similar products are available in the same outlet, the full difference in price between the two is explained entirely and solely by the difference in quality of the two products*".

For instance, the price difference of Coca Cola in a glass bottle (of 25 cl), a can (of 33 cl) and plastic bottles of 0.5 and 2 litres is defined as a difference in quality. The standard unit prices of these different packages (say, price per litre) would differ as it not only depends on the contents, but also on the type of packaging and size of the unit. As long as the pricing item within each store / stall remains constant from period to period, the correct measurement of price changes will take place. Similarly, if two identical packages of a particular brand of an item are priced differently based on the location of production, these are different qualities of the same product, differentiated only by the 'Made in xxxx' indicator.

Two apparently different products representing the technologies for the same application over time may be treated as different qualities of the same product. This is a practical solution to the problem of re-assigning weights. Often, the weights set does not yet include weights for the (non-existent) new product. This would, for example, cover CD players replacing cassette players, DVD players replacing VCRs,

etc. Treatment of the different commodities as different qualities of the same general good (CD players and cassette players both provide music without radio, for instance) is that the introduction of new technology is not even, nor very fast, and it normally starts from the metropolis to the towns to the rural areas. At any one time, some items may not be available in some areas but may be the major technology in others.

A.3.3 Weights hierarchy

The hierarchy given in the weights file determines the calculation and aggregation procedures for a particular application. The weights file for an application should contain weights for all (sub)-aggregates for which indices have to be calculated for the application.

The hierarchical structure is thereby very important as it determines the index series. From Item-level to the overall aggregate for the country, eight levels of aggregation are available in the system.

A typical application may have the following levels:

- Overall (country) total
- Geographical region (province/state/...)
- Urban/rural area
- Food / non-food products
- Major groups
- Groups
- Item-level.

It should be noted that the system does not require identical structures for all levels. For instance, the capital area may be a separate province, but may consist only of an urban stratum. Similarly, some of the group levels in the food section and some non-food major groups may not be used in the item classification scheme.

A.4 Documentation of the PRIMA system

The documentation of the PRIMA system contains a set of five manuals, namely:

a) Installation manual

b) Desktop user manual (also available as an online help function)

c) PDA user manual

d) Indexer manual and

e) Analytical module manual.

A.4.1 Installation Manual

The installation manual provides details of the technical aspects of the installation of the system in a country. This includes the installation of:

a) MySQL server system

b) Dataset of the country / region

c) PDAs and their links to the desktop and/or server

d) Desktop application

e) Indexer application

f) Analytical module and

g) Supporting software tools.

This manual also provides a brief description of the technical specifications and requirements of the PRIMA system. It is primarily intended for use by the IT staff of the implementing organization. Extensive use is made of screenshots showing the texts that appear on the computer screen while carrying out the various tasks. This reduces the chance for errors, increases the speed of installation and facilitates troubleshooting if problems occur.

A.4.2 Desktop User Manual

This manual describes the basic methodology of the PRIMA system. Furthermore, it describes all the processes used in the management functions included in the desktop for the maintenance and updating of the CPI system. These include:

a) Preparation of data entry forms (printed versions or electronic ones used on PDAs);

b) Loading data from PDA to database;

c) Manual data entry, review and corrections on the desktop;

d) Maintenance of the indexer framework (uploading weights sets, maintaining database);

e) Revision to the list of markets surveyed;

f) Revisions to the list of items, qualities and shops;

g) Revision of lists of common items;

h) Review and correction of data with large discrepancies;

i) Copying pricing information for common items to all markets;

j) Checking price reporting status for a period;

k) User access maintenance; and

l) Various other system maintenance functions as added from time to time.

A.4.3 PDA User Manual

This manual describes the use of the Personal Digital Assistant (PDA) for data capture in the PRIMA system. Data capture consists of three separate operations, namely:

a) Data entry

b) Import of the empty questionnaire form and

c) Export of the filled questionnaire form.

As data entry is the most common operation, it appears at the top of the list, although the second operation, the importation of a blank questionnaire form, needs to be executed before data entry can be started. In the PRIMA system, all data entry is organized by market or store, depending on the situation in the country. In African countries, it is generally by market, and the data entry form for one market is loaded at a time.

The questionnaire forms for a market (or store) are pre-filled with the latest available prices for that market or store, so that a price comparison can be carried out immediately in the markets. In cases where prices have changed by more than the pre-set range (25 percent in general), data entry can only continue after the pricing agent has given the reason for the major change. This information is stored with the prices and transferred to the database where it is made available for review in the discrepant data supervisor option.

The PDA manual comes in three options. The latest one is for the Android system, an older one for the Windows mobile operating system and the last one for the outdated Palm operating system. While the functionality of all versions of the programme is the same, the screen display features of the operating systems differ and therefore require separate descriptions.

A.4.4 Indexer User Manual

This manual describes the various options available in the indexer module. In addition to the step-by-step instructions on how to use the indexer, it also includes a brief overview of the indexing options and methodology available in the PRIMA system. The maintenance of the weights for the various index sets is documented in the Desktop User Manual and is therefore not included in this manual.

A.4.5 Analytical Module Manual

This manual details the range of options available for the analysis of the price indices. A continuously increasing range of outputs is available. These are either produced as html files, standard tab/comma-delimited files and jpg-format graphs for easy use in spreadsheets and reports.

While most of the analysis is conducted on single datasets, where appropriate, two datasets may be linked, for instance, for the generation of Fisher indices. The available data can be analysed at any level of aggregation from the item to the overall total. At each level, filters can be used so that any combination of items can be grouped in specialized indices. The aggregated values can be re-calculated in the process using the relative weights of the items selected.