TECHNICAL ASSISTANCE REPORT

UKRAINE

Report on Consumer Price Index Mission

(March 25–29, 2024)

JULY 2024

Prepared By
Claude Lamboray

Authoring Department:

Statistics Department

DISCLAIMER

"The contents of this document constitute technical advice provided by the staff of the International Monetary Fund to the authorities of Ukraine (the "CD recipient") in response to their request for technical assistance. This document (in whole or in part) or summaries thereof may be disclosed by the IMF to the IMF Executive Director for Ukraine, to other IMF Executive Directors and members of their staff, as well as to other agencies or instrumentalities of the CD recipient, and upon their request, to World Bank staff, and to other technical assistance providers, unless the CD recipient specifically objects to such disclosure (see Operational Guidance for the Dissemination of Capacity Development Information). Publication or Disclosure of this report (in whole or in part) to parties outside the IMF other than agencies or instrumentalities of the CD recipient, World Bank staff, and other technical assistance providers shall require the explicit consent of the CD recipient and the IMF's Statistics department."

Table of Contents

Acronyms and Abbreviations

COICOP	Classification of Individual Consumption by Purpose
CPI	Consumer Price Index
IQR	Interquartile Range Method
ML	Machine Learning
SSSU	State Statistics Service of Ukraine
STS	State Tax Service of Ukraine
TA	Technical Assistance

Summary of Mission Outcomes and Priority Recommendations

1. A technical assistance (TA) mission was conducted in Vienna, Austria to assist the State Statistics Service of Ukraine (SSSU) with introducing administrative scanner data for the compilation of the consumer price index (CPI). The SSSU plans to use administrative scanner data from the State Tax Service (STS). This data source covers the transactions recorded by cash register machines that retailers have been legally required to use. The mission provided advice on the main processes for using this data source in the Ukrainian CPI.

2. Incorporating scanner data will improve the quality of the CPI and reduce field price collection. The regional, product, and temporal coverages will be improved once these data are introduced. Weighted price indices can be compiled as both price and revenue information are available in the scanner data.

3. Some pre-processing will be required by the STS before transmitting the data to the SSSU. The STS data are big data, containing information on each individual transaction made by a retailer for an item. The mission recommended combining transactions referring to the same item sold in all the outlets of a region during a given month. As a formal item code is not available, the text label must be used for identifying transactions of the same item. An average price is then obtained by dividing the total revenue by the total number of units sold. Some data cleaning is required before calculating these average prices. The data must also be connected to the classification used in the CPI. The SSSU has already developed string-searching algorithms that identify patterns in the label of the items so that they can be automatically assigned to a CPI category.

4. The SSSU should agree with the STS on the characteristics of the data that will be made available. One should specify the pre-processing calculations, the variables included in the data, and the frequency and timeliness of the data supply, including the provision of back data for the first transmission. The data supply should be formalized by updating the protocol in the memorandum of understanding between the SSSU and the STS. In accordance with this agreement, the generation, transmission, reception and storage of the data will have to be prepared from a technical point of view.

5. The SSSU must then further process the data before it can be used for index calculations. The mission recommended linking the items in the scanner data to the lowest level of the CPI hierarchy (i.e. the level of the 335 representative products). The SSSU should further review and update the classification suggestion made by the string-searching algorithm. Processes and tools should be established to support the classification of items. Finally, outlier methods should be implemented in order to detect and remove items with unusual prices.

6. Experimental price indices should be calculated for a selection of products using the fixed basket method. This method is closer to the current practice, less data demanding and easier to interpret than more advanced index methods. Indices should be calculated only for a limited selection of products (for example some food products or pharmaceuticals) despite the potential wide scope of the STS data. The indices should be calculated at the level of the representative product by region. This is the same level at which elementary price indices are obtained from the manually collected prices. Such an approach will facilitate consistent comparisons with the current price series and the future integration of the scanner data into the CPI.

7. To support progress in the above work areas, the mission recommended a detailed action plan with the following priority recommendations.

TABLE 1. Priority Recommendations

Target Date	Priority Recommendation	Responsible Institution
August 1st 2024	Arrange regular access to the administrative scanner data	SSSU
December 31st 2024	Calculate experimental price indices from the administrative scanner data	SSSU

Further details on the priority recommendations and the related actions/milestones can be found in the action plan under Detailed Technical Assessment and Recommendations.

Section I. Detailed Technical Assessment and Recommendations

TABLE 2. Action plan

Priority	Action/Milestone	Target Completion Date
Topic: Data access and data preparation		
High	Agree with the STS on the structure of the data to be transmitted to the SSSU and sign the additional agreement	May 15st 2024
High	Prepare the data flow and transmission protocol according to the agreement	June 1st 2024
High	Start accessing the data on a monthly basis	July 1st 2024
High	Confirm the quality of the received data files	August 1st 2024
Topic: Classification		
High	Specify pattern matching rules for classification	Completed
Medium	Discuss with the STS about availability of additional information on the classification of the items	September 1st 2024
Medium	Build a reference table for archiving the link between items and COICOP categories	September 1st 2024
High	Review and update pre-classified data for the products that are in-scope of the pilot study	September 1st 2024
Topic: Data cleaning		
Medium	Discuss with the STS on applying the selected outlier methods during the initial transformations of the raw data.	May 1st 2024
Medium	Decide on the thresholds for the outlier and dumping filters and implement these rules on the data	October 1st 2024
Topic: Index compilation		
High	Draw samples for the initial basket	October 1st 2024

Priority	Action/Milestone	Target Completion Date
High	Decide on the principles for the treatment of replacements	October 1st 2024
Medium	Start preparing the IT tools for index compilation	November 1st 2024
High	Calculate experimental price indices for a selection of products	December 31st 2024
High	Compare the results to the current price series and adjust the methodology if needed	December 31st 2024
Medium	Update the methodological document for the pilot study	February 1st 2025

A. INTRODUCTION

8. The SSSU plans to use administrative scanner data from the STS for the purpose of calculating the CPI. This administrative data source covers the transactions recorded by cash register machines that retailers have been legally required (with some exceptions) to use. The STS would transmit these data in a pre-aggregated form to the SSSU for CPI calculations.

9. The administrative scanner data could improve the product, geographic, and temporal coverage of the CPI. Scanner data provide improved product coverage as more items sold by a retailer can be included, instead of sampling only a few varieties. The geographic coverage is also improved, as more outlets can be included. The time coverage can be expanded, as scanner data cover a large part of the reference month for a specific item, as opposed to observing a price in an outlet once or twice per month. This will also help to better capture sales, promotions, and discounts.

10. Scanner data will improve the accuracy of the price indices. Scanner data contain information not only on prices but also on the number of units sold. Such detailed weighting at the individual item level makes it possible to sample items according to their importance and to apply weighted index formulas.

11. A multi-source data collection approach will reduce price collection costs and make the CPI production system more resilient. Field price collection is the main data source for the CPI of Ukraine. The administrative scanner data could eventually replace part of the manual field price collection that is currently conducted by staff from the regional bodies of the SSSU.

12. The work on scanner data in the CPI will span over several years. In 2024, efforts should be made to secure regular access to the data and to calculate experimental price indices. These experimental indices can then be compared to the current series. An action plan (Table 2) has been prepared with the main steps to be undertaken for conducting such a pilot study. Further TA may be needed once progress has been made on the pilot study. Depending on the outcome of the pilot study, the CPI production systems could be adjusted during 2025 for an integration in the CPI by January 2026. The integration of scanner data in a CPI is typically conducted in an incremental manner. For example, one could first integrate food products and pharmaceuticals before gradually extending the scope to other products.

13. The use of scanner data will change the production process of the CPI. In the future, the processing of scanner data will create new tasks at the central level while reducing the data collection activities conducted at the regional level. The main processes for using scanner data in a CPI are: (i) data

collection and data preparation; (ii) classification; (iii) data validation; and (iv) index compilation and data integration. These processes are further discussed in this report.

B. DATA COLLECTION AND DATA PREPARATION

14. The STS data are big data. The data source includes prices data on each individual item that was transacted between a retailer and a buyer. During the mission some sample files were examined. One day of transactions amounts to over 77 million lines, with each line representing a transaction for a specific item. Currently data are collected from over 900,000 cash register machines, covering different types of outlets across the country for a wide range of products. The coverage of this data source is likely to further increase in the future.

15. Some initial transformations should be done by the STS before the data are transmitted to the SSSU. Some pre-aggregation will help to reduce the size of the data**.** It will be more practical for the SSSU to process smaller data sets. Moreover, some degree of aggregation avoids the disclosure of transaction level data that may be considered confidential.

16. The mission provided advice on the level at which average prices should be calculated. The following dimensions must be taken into account when calculating an average price with scanner data: (i) the time dimension; (ii) the outlet/regional dimension; and (iii) the product dimension. Conceptually, transactions can be combined into a single average price to the extent they are of the same quality.

- **Time dimension.** In general, the manual price collection is conducted between the 1st and the 25th day of each month. The average price should align with the temporal coverage of the current CPI. The data are available at the STS with a delay of approximately 5 days. In a production environment, a data supply at the end of the month would leave enough time for processing and analyzing the data in view of publishing the CPI on the 9th day of the following month. The choice of the temporal coverage should be confirmed following the results of the pilot study and discussed with members represented in the CPI advisory board.
- **Outlet/regional dimension.** The regional stratification of the CPI consists of 25 regions. The average price should be calculated according to the same regions. Such an approach will facilitate integration with the CPI as prices data will become available for each regional stratum. Compared to the current practice, the geographic coverage of the prices data will improve. The data entering the average prices should refer to transactions of an entire region, and not only to transactions that take place in the current CPI price collection cities. There could be a possible unit value bias[1] in the average prices because transactions from different types of outlets (different chains, small and big outlets) are combined. However, the supply of data by outlet or outlet-type is likely not to be possible because of confidentiality.
- **Product dimension**. Barcodes are not included in the STS data because the supply of such information by the retailer is not compulsory. In the absence of a formal item identifier, the label (text string) must be used to identify the same items. Some text string cleaning should be conducted in order to group together identical items that have slightly different labels.

[1] At the same time, this approach reduces the “outlet rotation bias” (see paragraph 2.107 in the CPI Manual CPI Manual - Concepts and Methods (2021)) and the “outlet substitution bias” (see paragraph 2.125 in the CPI Manual CPI Manual - Concepts and Methods (2021)) that can be encountered in a CPI.

17. Average prices should be calculated as unit values. A unit value is obtained by dividing the total revenue by the total number of units sold for all the transactions that fall within the specification of the average price. This average price (unit value) is not identical to the arithmetic average of the prices for each transaction line in the raw data. This is because more than one unit may be sold in one transaction line.

18. The data supplied to the SSSU should include the variables that are needed for CPI processing. Ideally, the following variables should be made available for a given item in a given region in a given month:

- Date (month)
- Label describing the item (text string)
- Region code
- Total revenue (for a given item/region/month combination)
- Total number of units sold (for a given item/region/month combination)
- An average price (obtained as total revenue divided by total number of units sold)
- COICOP category (see section C in this report for further details)
- Number of transaction lines in the raw data underlying the calculations of the average price (for quality control)

19. It remains uncertain if data on revenue and number of units sold can be supplied because of the confidentiality of these variables. In general, scanner data used in CPI calculations contain both prices and number of units sold. If access to revenue or number of units sold cannot be arranged, it would be beneficial for the purpose of CPI calculations to provide an alternative indicator that measures the "importance" of an item. For example, the revenue could be transformed into a revenue share by category. Another option would be to rank the revenue in decreasing order, and to transmit the rank of an item in its category instead of the revenue variable itself.

20. The memorandum of understanding between the STS and the SSSU should be updated by including a dedicated protocol for the supply of scanner data. Such an agreement should define the variables included in the data set, the rules of the data transmission (frequency and timeliness), including the supply of historical data for the first transmission, and other provisions such as any pre-processing to be conducted on the data by the STS. An agreement ensures continuity in the supplied data and helps to avoid uncoordinated changes or disruptions in the transmission of the data.

21. **The SSSU should conduct some initial quality checks after reception of the data files.** This initial validation could include verifying the size of the file, checking the variables that are included in the file, and compiling some basic summary statistics for the main variables.

Recommended Actions:

- The SSSU to agree with the STS on the data to be transmitted to the SSSU and to formalize the data supply by updating the memorandum of understanding (MOU). Data variables, frequency, timeliness, and STS pre-processing should all be addressed in the updated MOU.
- The SSSU to prepare for the reception of the data and to agree with the STS on a data transmission protocol.

- The SSSU to start receiving the data on a regular basis and to confirm that the data meet the agreed requirements.

C. CLASSIFICATION

22. The items included in the scanner data must be connected to the classification used in the CPI. The Ukrainian CPI uses the Classification of Individual Consumption by Purpose (COICOP) which is broken down into 335 representative products at the lowest level (referred to as the 7-digit COICOP level). The mission recommended to link the items in the scanner data to the level of the representative products (7-digit COICOP). This will make it possible to calculate price series with the same level of detail as in the current CPI. The classification process is a crucial step to ensure the accuracy of the indices.

23. The SSSU has already developed a string-searching algorithm that identifies patterns in the labels of the items. This method automatically links an item to a 5-digit or even to a 7-digit COICOP category, depending on the product. While this method is not free from errors, it helps to organize, to some extent, the data according to COICOP. The text mining rules must continue to be maintained, as the patterns in item labels can change over time.

24. The mission recommended to review and possibly update the classification suggestion made by the string-searching algorithm. Pattern matching is not perfect and can lead to classification errors. Moreover, 7-digit COICOP categories must still be selected for items that have only been classified by the method at a higher level. In the short term, this review can be done manually. The classification burden could be kept to a minimum by treating only items that are in-scope of the pilot project and by focusing on the most sold items.

25. Processes and tools should be established to support the classification of items. The item-COICOP relation should be stored in a separate reference table. This reference table can then be merged with the monthly scanner data files in order to obtain the corresponding COICOP category for each item. Such a strategy will make the classification process efficient, as every item has to be treated only once, when it enters the reference table. It will ensure that the classification is consistent across regions (i.e. the same item sold in different regions is assigned to the same COICOP category), and across time (i.e. the category of an item does not change over time). In the initialization phase, a lot of the data must be processed at once. In the following months, the reference table is only augmented with the items that have not yet been classified. In general, the recurrent phase of monthly updates will be less resource demanding than the initial phase.

26. The SSSU should continue to explore other methods to further optimize the classification of the data. Subject to the availability of another classification to which the scanner data are mapped, a correspondence table could be built between the categories of this auxiliary classification and COICOP. It would then be possible to automatically assign a COICOP category to an item by bridging via this auxiliary classification. Machine Learning (ML) methods can also be used for automatic coding of the data. ML methods create classifiers that learn from already classified items. Once a classifier is estimated, any new incoming items can be automatically classified. A fully classified training data set must be available to estimate a classifier and evaluate its performance.

Recommended Actions:

- The SSSU to specify pattern matching rules in order to pre-classify the scanner data as much as possible.

- The SSSU to create a reference table for maintaining the COICOP-item relation.
- The SSSU to additionally review the classification proposed by the pattern matching method for the items that are in-scope of the pilot study.
- The SSSU to explore other methods that could further improve the classification of the data.

D. DATA CLEANING

27. Outlier detection methods can be used to identify and remove unusual observations. Such observations could introduce additional volatility or even some bias in the final price indices. The data cleaning should be conducted at two levels: data se.

- At the level of an individual transaction line (raw data)
- At the level at which average prices are calculated (pre-aggregated data)

28. One could remove transactions with close-to-zero prices in the raw data. This approach is sensible to the extent that these prices can be errors. Such transactions could also be linked to some discount scheme, in which case they should not be considered as outliers. From a conceptual point of view, reduced prices should be covered in the CPI and should not be removed during data cleaning.

29. The Interquartile range (IQR) method can be applied on the raw data to identify transactions with unusual prices. The IQR method sets upper and lower bounds derived from the price distribution of a set of similar transactions (for example transactions with the same item sold on the same day). A transaction is considered to be an outlier if its price lies above or below a certain threshold defined with the help of the third and first quartiles of the price distribution**:**

$$Transction\ \ price < Q1 - 1.5 * (Q3 - Q1) \quad or \quad Transction\ \ price > Q3 + 1.5 * (Q3 - Q1)$$

The SSSU has already made some tests with the IQR method. The results shown to the mission confirmed the applicability of this method in this context.

30. The IQR method could also be applied with respect to the number of units that are sold in a transaction. Some lines in the raw data exhibit a large number of units sold in a single transaction. According to the SSSU, these lines could refer to small wholesale transactions, to bulk registrations by the retailer, or to bulk purchases. It may be prudent to exclude such transactions on conceptual or on practical grounds.

31. Outlier detection should also be applied at the level at which average prices are calculated. The average prices have a direct impact on the final price indices. An item in a given month is considered to be an outlier if the change in the average price compared to the previous month is above or below a certain threshold. This outlier method thus depends on price relatives, and not on price levels.

$$\frac{p_i^t}{p_i^{t-1}} < l \qquad or \qquad \frac{p_i^t}{p_i^{t-1}} > u$$

Some empirical tests should be conducted to specify these thresholds. Typical values for an outlier filter in scanner data applications are l=0.25 and u=4**.**

32. A dumping filter should be implemented in order to avoid a potential downward bias on the index. Items sometimes exit the market with a very low price. Including such end-of-lifecycle prices can introduce a downward bias into the price index. This is because the large price decrease caused by

the last price observation will not be offset by a price increase in the subsequent months. This causes a persistent downward impact on the price index. In practice, a dumping filter flags observation if both price and quantity changes are "low"[2].

$$\frac{p_i^t}{p_i^{t-1}} < d_p \qquad and \qquad \frac{q_i^t}{q_i^{t-1}} < d_q$$

Some empirical tests should be conducted to specify these thresholds. Typical values for a dumping filter in scanner data applications are d_p=0.7 and d_q=0.3.

Recommended Actions:

- The SSSU to discuss with the STS applying the selected outlier methods during the initial transformations of the raw data.
- The SSSU to decide on the thresholds for the outlier and dumping filters and to flag the data accordingly before the data enter the index compilations.

E. INDEX COMPILATION AND DATA INTEGRATION

33. There are different methods that can be used to obtain price indices from scanner data. Typically, these methods must address two key features of scanner data. First, the set of available items in a region is not static, but changes from one month to another. Second, the method should take into account not only the prices but also the economic importance of the items.

34. The fixed basket method [3] relies on a representative set of items that is selected in the base period. The initial "basket" should be selected with the help of formal sampling methods. For example, revenue shares[4] can be used to determine the significance of an item in its category. Items are then selected for inclusion in the basket through cutoff sampling (see Table 3 for an illustration of this method). Cutoff sampling is especially relevant if the revenue data are highly skewed. The idea is to select the most important items and disregard the many items that contribute little sales. The revenue shares for selecting the basket could be based on data from several months, and not only from the base month. This will help to reduce the impact of seasonality on the selection of the basket.

TABLE 3. Selection of items with cutoff sampling (cutoff threshold at 50 Percent)

Item	Revenue	Units sold	Price (=revenue /Units sold)	Revenue shares	Cumulative shares	Sample
Item 1	18 756	598	31.36	18.1%	18.1%	1
Item 2	17 398	349	49.85	16.8%	34.8%	1
Item 3	16 498	165	99.99	15.9%	50.7%	1

[2] This pattern is at odds with the standard pattern of price decreases (due to promotions for example) that come with spiking quantities.

[3] See the following paper for an implementation of the fixed basket method: "Implementing scanner data in the Danish CPI", paper presented at the 2014 Meeting of the UN Group of Experts on Consumer Price Indices, available at https://unece.org/sites/default/files/2021-04/WS4_05_Implementing_scanner_data_in_the_Danish_CPI.pdf.

[4] Either revenue or number of units sold can be used for sampling. Conceptually, the choice between these two variables depends on the price index formula that is eventually used. Items should be sampled based on revenue if using a Jevons index (see paragraph 4.19 in in the CPI Manual - Concepts and Methods (2021)).

Item 4	15 449	765	20.19	14.9%	65.6%	0
Item 5	15 327	438	34.99	14.8%	80.4%	0
Item 6	8 754	387	22.62	8.4%	88.8%	0
Item 7	6 594	267	24.70	6.4%	95.1%	0
Item 8	3 128	298	10.50	3.0%	98.1%	0
Item 9	1 065	132	8.07	1.0%	99.2%	0
Item 10	856	54	15.85	0.8%	100.0%	0

35. In the fixed basket method, the same set of items is priced in each month. However, some of the initially selected items may not be available in later periods. In line with standard CPI practice, one must select a replacement item that is then introduced into the basket. In particular, the following must be done in month t:

- Access the basket of the previous month t-1.
- For each item of the t-1 basket, check if the item is still available (or representative) in month t.
    - If YES: add the item to basket of month t.
    - If NO: select a replacement item to be added to the basket of month t.
- Calculate a Jevons index (geometric mean of price relatives) for the month t basket.

36. With a fixed basket method, some efforts should be spent on the treatment of replacements. The price of a missing item can be imputed for one or two months before a replacement item is selected. If an item is considered to be permanently missing or if it is found to be not representative anymore, a new item must be selected from the scanner data and linked into the basket. The selection of a replacement item can be done manually, or it can be automatized to some extent by specifying some rules. For example, one could select the best-selling item in the current month that is not yet included in the basket. There are different methods that can be considered for introducing the replacement item into the basket:

- Overlap pricing, as the price of the replacement item in the previous month is often available in the scanner data.
- Quantity adjustment, if there is a change in the content of an item.
- Direct comparison, for comparable replacements
- Mean imputation, for non-comparable replacements

37. The fixed basket method favors comparability over representativity. Comparability is achieved by pricing the same set of items over time and by minimizing the need for replacements. However, the representativity of the basket selected for the base period may deteriorate over time. That is why this method should best be combined with a regular (e.g. annual) resampling schedule of the items that are included in the basket.

38. Scanner data makes it possible to apply weighted index compilation methods. Ideally, with scanner data, we aim to use all the items (not only a sample) in the index calculations. Moreover, each item should be weighted according to its importance. The Törnqvist price index is one of the preferred index number formulas. It is consistent with the economic theory of price indices and it has good

axiomatic properties. It can be seen as a weighted version of the Jevons index, where the weights correspond to the average revenue shares of the items in the two comparison periods.

$$I^{t-1,t} = \prod_i \left(\frac{p_i^t}{p_i^{t-1}} \right)^{0.5*(s_i^t + s_i^{t-1})}$$

39. Monthly chaining of Törnqvist indices leads to chain drift[5] and is not recommended. In practice, the price indices are calculated based on the set of items that are available in the two comparison periods (matched set of items). In order to maximize the potential overlap of items, it would be natural to calculate Törnqvist indices between two consecutive months, and then chain the resulting period-to-period Törnqvist indices. Unfortunately, such a strategy could lead to chain drift and should therefore not be implemented as such. The chain drift problem can be solved by using the Törnqvist index as building blocks of a multilateral method[6].

40. The mission provided initial training on these index compilation methods. Excel files were provided to the SSSU to illustrate the different methods. Some preliminary index calculations were also made on STS test data that were analyzed during the mission.

41. The mission concluded that the fixed basket method is the best option for the pilot study. A multilateral method would be preferable from a theoretical point of view. However, the fixed basket method is easier to implement because it is closer to current practice. The calculations can be made with the help of IT tools already available to the CPI team (e.g; Excel). The method is less data demanding as it can also be implemented with more limited information on classification and revenue. This method will facilitate the evaluation of the quality of the input data and the interpretation of the resulting indices.

42. To implement the fixed basket method, the sampling method for the initial basket and the method for treatment of replacements must be decided. The pilot study will help to identify the appropriate design of the fixed basket method. The resulting experimental price indices should be compared to the currently calculated series. The choice of the compilation method could be revisited after this initial pilot study.

43. In the pilot study, price indices should be calculated only for a limited selection of products despite the potential wide scope of the administrative scanner data. During the pilot phase, the CPI team will have to work both on the regular calculation of the CPI and on the experimental calculation with scanner data. The SSSU plans to focus on some food products (COICOP 01.1) and pharmaceutical products (COICOP 06.1.1). These products make up a significant share (more than 40 percent) in the CPI basket.

44. The Jevons and Törnqvist price indices satisfy the "commensurability test". This test[7] requires that the price index should not change if the unit of measurement is changed, for example, if the

[5] See Table 10.1 in the CPI Manual - Concepts and Methods (2021) for a numerical example of chain drift.

[6] More details on this method can be found in section 4.1 in "Guide on multilateral methods in the Harmonised Index on Consumer Prices (HICP) — 2022 edition", Eurostat, 2021, available at https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-21-020

[7] See paragraph 8.30 in the CPI Manual - Concepts and Methods (2021)

prices are expressed per liter rather than per pint. As a consequence, the prices of the items do not need to be transformed into a common unit of measurement for calculating these price indices.

45. In addition to the elementary price indices that enter the CPI, there is a user demand for average prices. A common unit of measurement would be needed in order to calculate meaningful average prices. Information on the unit of measurement is not available in a standardized manner in the scanner data. Instead, it must be extracted from the labels describing the item. Some labels do not include any information on the unit of measurement. It has to be further investigated how the scanner data can best be leveraged to compile such average prices.

46. Scanner data indices should be calculated at the level of the representative product by region. This is the same level at which elementary price indices are currently calculated by the regional offices from the manually collected prices. Such an approach will allow a consistent comparison with the current price series and facilitate the integration of the scanner data into the CPI. In the future, scanner data indices can be progressively integrated in the CPI for selected strata (defined as a representative product in a region). As the scanner data in principle cover most outlets in a region, there should be no need for a parallel field price collection in a given stratum. Ideally, a stratum could be based either fully on field price collection or fully on scanner data. The SSSU can continue to use the same annually updated CPI expenditure weights for combining the stratum indices. In the end, the CPI will be based on different data sources (field price collection and scanner data) and on different index methods applicable to each data source.

Recommended Actions:

- The SSSU to decide on a specific sampling and substitution methodology for the fixed basket method.
- The SSSU to calculate experimental prices indices for a selection of products using the fixed basket method.
- The SSSU to compare the experimental price indices to the current price series and adjust the scanner data methodology if needed.
- The SSSU to update the methodological document describing the pilot study.

F. OFFICIALS MET DURING THE MISSION

Name	Position
Ms. Nadiia Profatska	Deputy Director-Head of Division for Consumer Price Statistics, Department for Price Statistics, SSSU
Ms. Anastasiia Simson	Senior Specialist of the Division for Consumer Price Statistics, Department for Price Statistics, SSSU
Mr. Pavlo Polikarchuk	IT and Data expert
Ms. Mariia Daniuk	IT and Data expert