Opportunities for integration of scanner data into the Belarusian CPI

K.Hrykhanava and A.Yarkovets

National Statistical Committee of the Republic of Belarus

1. Introduction

Consumer price index in the Republic of Belarus is a Lowe index, based on monthly/weekly price collection by 88 collectors, in 31 cities, more than 8500 outlets and organizations. Price collection is carried out in small, middle shops, large outlets and supermarket chains, in the markets. Most prices are collected manually by price collectors.

To explore the possibility of the scanner data use as alternative source for the Belarusian CPI, repeated requests have been made to the main retail chains to negotiate the provision of data to statistical bodies over the last ten years. The refusal of the retailers to provide the data was explained by the commercial secret, anti-competitor protection and exclusive rights to the original software.

But it has now become possible for Statistics Belarus to receive the data from the system, which collects cashier`s checks data nationwide for the tax bodies purposes (administrative data type). This system is intended for receiving, recording and processing of sales cash receipts by cash equipment registered in the system.

In this regard it's planned for the coming years to test the quality of the scanner data from the system, to make experimental calculations of the CPIs on sampled items from selected retailers.

2. The source of scanner data

The system which is considered as a source of scanner data in the Belarusian CPI, accumulates the data from payment documents (cashier`s checks) from retail outlets with over 200 $m^2$ trading area. Payment document formed on the cash register is transmitted to the pool of communication servers which carry out its format-logical control, pre-processing and record to the operational database. It`s actually raw data – unprocessed or minimally processed data which are unusable for quick analysis due to their lack of the unified structure or the unified way of accessing them. So that the data is suitable for analysis and calculation it was decided to integrate it with the Pool of electronic passports of trade items (ePASS).

ePASS is a centralized resource containing description of products according to the international standards of e-commerce. Manufactures and importers record their products in ePASS giving them detailed description in electronic passport. These descriptions are further transmitted to supply chains participants for application in

business process automation systems. Presently ePASS contains more than 13 million electronic passports of products.

Raw data:	ePASS:
Cash register number Payment document number **GTIN(EAN-13 mostly)** Date, time of issue Product quantity Unit price Discount	**GTIN** Package Functional product name Sub brand Product form Quantity per pack Weight Country of assembly code Country of manufacture code Country code where the product was last processed Code by National Classification of Industrial and agricultural production

3. Main problems, identified in the analysis

a.   There is list of goods sold which are not the subject to product numbering and barcoding. So nowadays there are issues with barcoding and identification of goods of variable quantities (e.g. loose confectionery, cheese, sausage that is sold by the weight, fruits and vegetables etc.). In this case the items sold are packaged directly by the outlet and have a local codes of limited circulation, assigned at each of them. These codes of limited circulation are not in the ePASS system. Eventually, even if the information on the transactions with goods of variable quantities is contained within the system of cashier`s equipment control, it`s impossible to identify the details required for the CPI purposes, such as GTIN, quantity sold, product identity.

b.   Goods, imported into the territory of Belarus, are not for the most part the subject to compulsory registration in ePASS (excluding special list of goods), therefore significant part of imported goods is lost for the scanner data.

c.   In the process of registering goods in the ePASS producers and importers fill in many descriptive features of the product, and often make mistakes in doing so. Analysis of a number of electronic passports of trade items revealed that the main mistakes are misclassification according to the National Classifier of Industrial and

Agricultural Production and insufficient or very short filling in of the goods characteristics on the electronic passport.

d.   Lack of an appropriate classifier to be included in ePASS system that allows grouping and qualitative sampling at an elementary level instead of the National Classifier of Industrial and Agricultural Production, which is too aggregated.

4. Actions planned to integrate scanner data in the CPI

The System, accumulating cashier`s checks data may provide information on major outlets only (200 m2 trading area). But there is a large number of stores with a smaller sales area, located nationwide where people tend to buy (especially in the regions). So scanner data may be used only as additional source of price data for now.

Thus, in the first phase it is planned to work with three product groups: milk, cereals and soft drinks. There are two ways considered. Since GTINS, belonging to these groups can be identified only on the basis of the Classifier of Industrial and Agricultural Production (highly aggregated levels), the option of preliminary implementation of GTINs sample and price observation for selected GTINs during the year is considered.

The second choice for use of scanner data assumes an initial selection of GTINs according to the Classifier of Industrial and Agricultural Production, and subsequent separation of the sample by homogenous products. This work is planned for the five largest retail chains in the capital city for 2022 and 2023.

At the moment there are many questions and problems for the use of scanner data in Belarusian CPI related both to the quality of the data and the method of calculation choice. There is also a technical problem of processing big data, choosing the way of data transfer. Belarusian CPI stuff initiated numerous discussions on the quality of electronic passports improvement and on the possibility of weight goods labeling.

For now scanner data in Belarusian CPI is considered as additional source to manual price collection.