



# GEORGIA

## TECHNICAL ASSISTANCE REPORT—SCANNER DATA FOR CPI

November 2021

This Technical Assistance report on Georgia was prepared by a staff team of the International Monetary Fund. It is based on the information available at the time it was completed on July 2021.

Copies of this report are available to the public from

International Monetary Fund • Publication Services  
PO Box 92780 • Washington, D.C. 20090  
Telephone: (202) 623-7430 • Fax: (202) 623-7201  
E-mail: [publications@imf.org](mailto:publications@imf.org) Web: <http://www.imf.org>  
Price: \$18.00 per printed copy

**International Monetary Fund**  
**Washington, D.C.**



# TECHNICAL ASSISTANCE REPORT

## GEORGIA

Report on Scanner Data for CPI Mission  
July 12–16, 2021

**JULY 2021**

**PREPARED BY**

*Vanda Guerreiro and Jens Mehrhoff*



**MEMBERS**

Armenia, Azerbaijan, Georgia, Kazakhstan, Kyrgyz Republic, Mongolia, Tajikistan, Turkmenistan, Uzbekistan

**PARTNERS**

Asian Development Bank, China, Korea, Poland, Switzerland

## DISCLAIMER

---

*The contents of this report constitute technical advice provided by the staff of the International Monetary Fund (IMF) to the authorities of Georgia (the “TA recipient”) in response to their request for technical assistance. This report (in whole or in part) or summaries thereof may be disclosed by the IMF to IMF Executive Directors and members of their staff, as well as to other agencies or instrumentalities of the TA recipient, and upon their request, to World Bank staff, and other technical assistance providers and donors with legitimate interest, including members of the Steering Committee of CCAMTAC, unless the TA recipient specifically objects to such disclosure (see [Operational Guidelines for the Dissemination of Technical Assistance Information](#)). Publication or Disclosure of this report (in whole or in part) or summaries thereof to parties outside the IMF other than agencies or instrumentalities of the TA recipient, World Bank staff, other technical assistance providers and donors with legitimate interest, including members of the Steering Committee of CCAMTAC, shall require the explicit consent of the TA recipient and the IMF’s Statistics Department.*

## MEMBERS

---



Armenia



Azerbaijan



Georgia



Kazakhstan



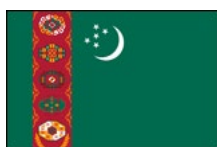
Kyrgyz Republic



Mongolia



Tajikistan



Turkmenistan



Uzbekistan

## PARTNERS

---



Ministry of Economy  
and Finance



Ministry  
of Foreign Affairs  
Republic of Poland



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs,  
Education and Research EAER

State Secretariat for Economic Affairs SECO

# CONTENTS

Glossary	4
<b>SUMMARY OF MISSION OUTCOMES AND PRIORITY RECOMMENDATIONS</b>	<b>5</b>
<b>DETAILED TECHNICAL ASSESSMENT AND RECOMMENDATIONS</b>	<b>7</b>
A. Introduction	7
B. SD Transmission and Coverage	8
C. Data Analysis and Management	10
D. Classification to COICOP	11
E. Officials Met During the Mission	12
<b>TABLE</b>	
1. Priority Recommendations	6

## Glossary

CCAMTAC	Caucasus, Central Asia, and Mongolia Regional Capacity Development Center
COICOP	Classification of Individual Consumption According to Purpose
CPI	Consumer Price Index
EA	Elementary Aggregate
EAN	European Article Number
Geostat	National Statistics Office of Georgia
MoU	Memorandum of Understanding
RPPI	Residential Property Price Index
SD	Scanner Data
TA	Technical Assistance

## SUMMARY OF MISSION OUTCOMES AND PRIORITY RECOMMENDATIONS

1. **The purpose of the mission was to assist the National Statistics Office of Georgia (Geostat) with continuing its modernization of the Consumer Price Index (CPI).** This was the first technical assistance (TA) mission to Georgia to specifically focus on scanner data (SD). Two earlier missions on the residential property price index (RPPI) (September 23–October 4, 2019, and November 30–December 4, 2020) already provided initial guidance and addressed the planned modernization of the CPI using SD. The mission was delivered remotely, and the sessions were recorded for future reference.
2. **Geostat aims at publishing the CPI with SD for food products and non-alcoholic as well as alcoholic beverages with data from two main supermarket chains by January 2023.** Experimental indices should be tested during 2022. The current mission addressed specifically the loading, inspection, and processing of the raw data files, and the mapping of products to a national breakdown of the Classification of Individual Consumption According to Purpose (COICOP) 1999. Another TA activity is planned towards the end of CY21. By then, the data transmission and the classification process should be finalized; and the follow up TA would assess the stability of the classification and identifying the next steps necessary to begin the compilation of experimental SD-based indexes.
3. **Meetings with the data providers are needed to secure data transmission, to agree on certain terms, and clarify data contents.** Currently there is no written agreement with the data providers. The mission provided examples of written agreements, such as a Memorandum of Understanding (MoU), for SD transmission. Several questions regarding content and data coverage need to be clarified. Additionally, since the supermarket product categorization is used in the classification process, Geostat should request the data providers to inform them in advance of any categorization changes, and moreover request the stability of the once agreed file structure.
4. **Negotiations with other supermarket chains should continue to improve the coverage.** Despite the limited market coverage, SD can still be introduced in the CPI compilation using turnover as weight to aggregate the supermarket indices produced from SD and those from price collection in the field. Nevertheless, the higher the coverage of the SD the better.
5. **While the Geostat staff has some prior knowledge in R, partially acquired during the RPPI missions, improving R programming skills should continue.** All data handling should be done in R since this allows for reproducibility and is less prone to errors. Spreadsheet software such as Excel is not capable of dealing with the amount of observations and performing the necessary processing steps. There are several online learning programs available for R, both internally and externally.
6. **Like the traditionally collected prices, SD should be classified to the national seven-digit level of COICOP (the elementary aggregate (EA) level).** This will help with the later

integration of SD in the CPI and blending it with traditionally collected prices. A correspondence table between the national classification and the European Article Number (EAN) codes was manually created and is maintained by Geostat. The correspondence table should rely on EANs and product descriptions in addition to the supermarket product categories and sub-categories. The consistency of the manual classification should be assessed for example by means of the number of EAs used for a given product category.

**7. The correspondence table created by Geostat for the classification needs further improvements.** When the supermarket data were linked with the classification file, there was an insufficient number of different products for each combination of city, EA and month to allow indexes to be calculated. Further, the classification file contains duplicate records, where the same product is attributed to different EAs. The results of these analyses were explained to Geostat and further research is needed before a follow-up on classification and index calculation could commence.

**8. To support progress in the above work areas, the mission recommended a detailed one-year action plan with the following priority recommendations carrying particular weight to make headway in improving the CPI.**

**Table 1. Priority Recommendations**

Target Date	Priority Recommendation	Responsible Institutions
<b>October 2021</b>	<i>Establish a firm written commitment (MoU) with the supermarket chains on the data deliveries at the management level.</i>	<b>Geostat</b>
<b>Continuous</b>	<i>Continue negotiations with other main supermarket chains on potential data deliveries.</i>	<b>Geostat</b>
<b>December 2021</b>	<i>Improve the quality of the manual classification and assess it vis-à-vis the raw data.</i>	<b>Geostat</b>

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

## DETAILED TECHNICAL ASSESSMENT AND RECOMMENDATIONS

Priority	Action/Milestone	Target Completion Date
<b>Outcome: Compile experimental SD-based indexes</b>		
L	CPI with SD for food products and beverages with data from two main supermarket chains is published.	January 2023
H	Plan for an increase of resources needed in the research and implementation phase.	November 2021
H	Establish an MoU with the supermarket chains on the content and format of the data deliveries, including automatization.	October 2021
H	Assess the consistency of the manual classification, particularly for new EANs.	December 2021
H	Check the validity and consistency of the manual classification and assess its quality.	December 2021
H	Analyze the limited product coverage of city-EA combinations after linking with the EAN-EA correspondence table, in terms of both the classification and the raw data.	January 2022
L	Receive data on a weekly or daily basis covering the full first two weeks of each month.	September 2021
H	Assess future developments in market structure to understand data coverage and representativity.	Continuous
L	Classify all products, even those out-of-scope, and keep separate conversion tables for each retailer.	December 2021
L	Continue discussions with other retailers on potential data deliveries.	Continuous
H	Process SD in R from the raw files to the final outputs.	September 2021
L	Participate in trainings on programming in R.	August 2021
M	Formalize the current classification as being at the seven-digit level.	January 2022
H	Classify the SD to that national seven-digit level of COICOP.	August 2021

### A. Introduction

**10. SD files are transmitted by the supermarket chains containing, for each individual item, the value of sales and the number of units sold in a city during a given period.** Instead of manually collecting the prices in the outlets, prices can be taken from SD files. Countries are rapidly expanding the use of SD for the compilation of the CPI. SD have several advantages over field price collection. Information on the actual expenditure for all items sold is available and this



can be used as “explicit,” monthly updated, weight at the item level. Price data are available over longer periods per item, it can better cover discounts, exclude items that are not actually sold, and the item sample is updated monthly.

**11. An increase in the quality of the CPI, and a reduction in CPI production costs and respondent burden are to be expected.** A better-quality CPI is expected due to the advantages described in the previous paragraph. In addition, in the long run, resource savings can be realized by reducing in-store price collection and data processing, since data are transmitted automatically, and all processing is further automatized. Nevertheless, the research and implementation phases require an investment namely in human resources skills and IT infrastructure.

***Recommended Action:***

- Plan for an increase of resources needed in the research and implementation phase, e.g., in order to achieve the development of a data analysis and management system (including the classification to COICOP).

## **B. SD Transmission and Coverage**

**12. This was the first mission to Georgia to specifically focus on SD.** Two earlier missions on the RPPI already provided initial guidance and addressed the planned modernization of the CPI using SD.

**13. In April 2020, the first set of SD was obtained from the two large supermarket chains.** First negotiations between Geostat and supermarket chains started at the end of 2019. Several face-to-face meetings were held to explain the purpose of using SD in the CPI. The restrictions to in-store price collection as a result of the pandemic enhanced cooperation since computer-assisted personal interviews were considered unsafe from both the Geostat and retailer’s side. Both chains have outlets in all the six cities surveyed in the CPI. Currently sales information is obtained mainly for food and non-alcoholic as well as alcoholic beverages.

**14. SD cover the first ten days of each month and are tentatively submitted to Geostat by the 15<sup>th</sup> of the month.** The data are transmitted in Excel format either via e-mail or WeTransfer.com. Currently, some months are missing, and data are often received with a delay. Ideally, data should be delivered automatically on a weekly or daily basis to mitigate the risk of failure of data transmission.

**15. Meetings with the data providers are needed to secure data transmission, to agree on the terms, and clarify data contents.** Currently there is no written agreement with the data providers. The mission provided examples of written agreements for SD transmission. Several questions regarding content and data coverage need to be clarified, for example:

- whether the duplicates of EANs on a given day and in a given city are due the data not being aggregated across outlets,
- why unit prices differ between outlets (if the above is confirmed) of the same supermarket chain, particularly in the same city,
- which types of discounts are included in the data and how, i.e., recorded to the usual EAN or given a separate code,
- if re-launches (the same product, typically in a new packaging, but with a different EAN) can be identified through stock-keeping units, and
- if information from the product descriptions can be broken up into brand, packaging size and so forth.

Since the supermarket product categorization is used in the classification of EANs to EAs, Geostat should additionally request the data providers to inform them in advance of any categorization changes. Moreover, it should be requested that the once agreed structure of the transmission files remains stable with regard to variables included, their naming, and the detail of aggregation provided, i.e., EANs, cities, days, outlets, and so forth.

**16. The Georgian retail market is characterized by a relatively low degree of concentration.** Information from business statistics reveals that the two supermarket chains cover 17 percent of retail trade turnover. A third chain that should be included in the future would bring that number up to 32 percent. This exemplifies the challenges in a competitive market of generating good coverage and representativity from the use of SD alone.

**17. Negotiations with other supermarket chains should continue to improve the coverage.** Despite the limited market coverage, SD can still be introduced in the CPI compilation using turnover as weight to aggregate the supermarket indices produced from SD and those from price collection in the field. Nevertheless, the higher the coverage of the SD the better.

***Recommended Actions:***

- Establish an MoU with the supermarket chains on the content and format of the data deliveries, including automatization.
- Receive data on a weekly or daily basis covering the full first two weeks of each month to balance the weekdays in the sample.
- Assess future developments in market structure to understand data coverage and representativity.
- Continue discussions with other retailers on potential data deliveries.

## C. Data Analysis and Management

**18. The mission provided extensive training in R by live coding the loading of the raw files and the below steps for inspecting the data.** The data obtained from the supermarket chains were filtered by regions and analyzed for errors, missing data and product categories that are out-of-scope. Thus far, these steps have been performed manually in Excel. All data handling should be done in R since this allows for reproducibility and is less prone to errors. Spreadsheet software such as Excel is not capable of dealing with the amount of observations and performing the necessary processing steps. The R code jointly produced with Geostat staff during the mission has been tested locally and is fully running; it is available for their use and further development. The R codes were prepared to perform the following steps:

- Checking the file structure
- Checking the size of the files
- Checking the number of items per period and city
- Checking for duplicate records
- Checking for missing values
- Checking for non-positive values in quantities sold or turnovers
- Checking the sum of turnover

**19. While the Geostat staff has some prior knowledge in R, partially acquired during the RPPI missions, improving programming skills should continue.** There are several online learning programs available for R, both internally and externally.

**20. The mission has made several steps for data analysis specifically for Geostat data.** After the initial inspection of the data, suspect data are dropped and the disaggregate information provided by the retailers is aggregated to chain-wide monthly totals. The individual months are then combined into a single file to prepare the further processing towards price indexes. The third step includes the linking of the retailer data with the EAN-EA conversion table and the restriction of the data set to in-scope products as well as to the six CPI cities. The mission uncovered that the classification of the raw data do not contain a sufficient number of different products for each combination of city, EA and month to allow indexes to be calculated. The results of this analysis were explained to Geostat and further research is needed before a follow-up TA on classification and index calculation could commence.

### ***Recommended Actions:***

- Process SD in R from the raw files to the final outputs.

- Participate in trainings on programming in R.
- Analyze the limited product coverage of city-EA combinations after linking with the EAN-EA correspondence table, in terms of both the classification and the raw data.

## D. Classification to COICOP

**21. The classification of the Georgian CPI is based on the COICOP 1999 and uses a further national breakdown to the six-digit level.** However, some six-digit aggregates consist of different items, making it effectively a seven-digit classification. The current seven-digit classification should be formalized, i.e., no six-digit level EA should be assigned different items. If there is no further breakdown of the six-digit level, the seventh digit should be zero, otherwise the seventh digit should be numbered consecutively. Like the traditionally collected prices, SD should be classified to the national seven-digit level of COICOP as this will help with the later integration of SD in the CPI and blending it with traditionally collected prices.

**22. All EANs are classified to the national six-digit level of COICOP.** A correspondence table between the six-digit national classification and the EAN codes was manually created and is maintained by Geostat. Every month new EANs that appear for the first time are again manually classified and added to the correspondence table. This process is supported by information on product categories used and supplied by the supermarket chains. However, the supermarket product categorization is not aligned to COICOP, and is thus, while helpful, not reliable. The classification should be based on EAN codes and product descriptions.

**23. Some supermarket product categories are distributed across a wide range of different EAs.** While there are some misclassifications by the supermarket chains and some product categories are rather generic, the consistency of the manual classification should be assessed for example by means of the number of EAs used for a given product category.

**24. In addition to the supermarket product categories, the product descriptions can be used for classification.** Particularly in the case of re-launches it may happen that the same product (description) has different EANs. In this case, the different EANs referring to that product description should be mapped onto one EA. Further steps include that when one supermarket product category is classified to one EA – assuming the categories are sufficiently stable and reliable – the results from manual classification can be transferred to a set of automatic rules. Should more detailed information become available or be extracted from the product descriptions, such as brand, this information could be used likewise. More advanced techniques using machine learning algorithms are not recommended at this juncture.

**25. Products that are out-of-scope are not classified and dropped, e.g., non-food items.** Further, SD contains products that are currently not covered by the CPI for they have a very low expenditure share in the household budget survey.

**26. Further to the above, the correspondence table between EAN and EA developed by Geostat needs some improvements.** There were duplicated records, where the same individual item as identified by the unique combination of EAN, product description, category, and sub-subcategory is attributed to multiple EAs. More generally, the correspondence table should rely more on EANs and product descriptions than on the supermarket product categories and sub-categories. The linking of the classification with the retailer data leads to a good portion of the data being deleted as being out-of-scope. But since the number of individual items in the linked data set falls short of the number of records in the correspondence table, the quality of the classification should be checked thoroughly.

**27. Geostat should keep a separate converter list for each retailer.** These lists should be consistent for the same combinations of EAN and description. In fact, the information from one retailer could be used to classify the other retailer if the products are the same. Finally, it is proposed to classify all products, even those that are out of scope to a dump category (e.g., 9999999). This would allow checking the functioning of the retailer-converter link, and non-food items, will still be dropped.

**Recommended Actions:**

- Formalize the current classification as being at the seven-digit level.
- Classify the SD to that national seven-digit level of COICOP.
- Check the validity and consistency of the manual classification and assess its quality.
- Assess the consistency of the manual classification, particularly for new EANs.
- Classify all products, even those out-of-scope, and keep separate conversion tables for each retailer.

## E. Officials Met During the Mission

Name	Institution
Gogita Todradze	Executive Director of Geostat
Giorgi Tetrauli	Head of Price Statistics Department, Geostat
Quji Bichia	Senior Specialist in the Consumer Price Statistics Division, Price Statistics Department, Geostat