Household Budget Survey within a new EU legal framework – towards higher quality and more harmonization, promoting innovative approaches.

Jakub Hrkal[1], Eniel Ninka[2], Erika TAIDRE[3], Teodora TCHIPEVA[4]

[1,2,4] Eurostat - Unit F.4 Income and living conditions; quality of life, Luxembourg

[3]Sogeti part of Capgemini, Luxembourg

Abstract

Since 1988, Eurostat has published statistics in the consumption domain every five years using data from the national household budget surveys (HBS). Up to the 2020 wave data have been obtained on a voluntary basis. Starting with the 2026 wave, HBS will be implemented under the Regulation (EU) 2019/1700 and related legislative acts. The new legal basis is expected to enhance the harmonisation of HBS data among European countries and with other EU social surveys. Moreover, it would bring improvements in various aspects of quality such as relevance, accuracy, timeliness, accessibility, and comparability.

Since 2017, Eurostat has supported European countries via various projects and exchanges of practices in improving the methodology and developing innovative and shareable solutions aiming at modernisation of HBS data collections. To this aim, Eurostat in collaboration with a task force on innovative tools and sources for the HBS, has supported the development, maintenance and testing of several innovative solutions consisting in webb and mobile app diaries including smart features like scanning and processing of shop receipts via optical character recognition and machine learning algorithms. As a result, fully fledged web and/or smartphone applications are available, and many countries plan to use them in the 2026 data collection wave either as the only mode or in a multimode setting. More recently, the focus of projects and grants has shifted towards the conceptualisation, proof of concept, and implementation of (trusted) smart surveys. The expected impact on data quality would be a reduction of recall and underreporting errors, real-time data checks and pre-validation performed in-app or in the back-office, and reduced post-fieldwork data processing time that would result in improved overall data quality and in terms of timeliness and accuracy, while maintaining data relevance and preserving privacy and confidentiality.

This paper first describes the potential for improvement of the quality of future HBS by comparing the current state of play and anticipated changes in view of implementing HBS under the new legislative framework. It also informs about the development of methodological and technical documentation for implementation of HBS and about a Eurostat project on the development of a new data production system, activities which will impact the quality of implementation, processing, and dissemination of HBS data. The paper then provides a brief overview on Eurostat activities aiming at supporting countries to improve the HBS methodology and further develop innovative and shareable solutions to modernise the HBS data collection.

Keywords: budget survey, innovation, shareable solution, data quality

1. Introduction

The Household Budget Survey (HBS) is a key instrument used to compile weightings for important macroeconomic indicators at EU level, such as the consumer price indices and the harmonised indices of consumer prices as measures of inflation, as well as for national accounts purposes. HBS provides detailed data on private households' total consumption expenditure linked to household characteristics such as income, housing and many other demographic and socioeconomic attributes. It therefore provides information on the economic and social living conditions of households and individuals in the Member States. The information from HBS is also used at EU level in the context of the consumer protection policy.

After depicting the state of play of the current HBS, the paper describes the potential for quality improvements of the future HBS at EU level based on three major trajectories. Firstly, by comparing the current state of play with the anticipated changes in view of the HBS implementation under a new legislative framework including the development of methodological and technical documentation. Secondly, the paper informs about a Eurostat project on the development of a new internal data production system. Finally, the paper offers a complete overview on the Eurostat activities aiming at supporting countries to improve the HBS methodology and further develop innovative and shareable solutions to modernise the HBS data collection.

Activities along these three development paths will significantly impact the quality of HBS implementation, including data collection and processing at national level, and transmission, processing and dissemination of HBS data and indicators at European level.

2. State of play of HBS data collection

Since 1988, Eurostat has published statistics in the consumption domain every five years using data from the national household budget surveys (HBS). Up to the last 2020 wave (which is still ongoing), seven data rounds have taken place with the following reference years:

- 1988, with the participation of 10 Member States (MS);
- 1994, 15 MS and 3 European Free Trade Association (EFTA) countries;
- 1999, 15 MS and 1 EFTA country;
- 2005, 25 MS, 1 EFTA country, and 5 candidate countries;
- 2010, 27 MS, 1 EFTA country, and 4 candidate countries;
- 2015, 28 MS, 1 EFTA country, and 5 candidate countries;
- 2020, 26 MS, 1 EFTA country, and 5 candidate countries.

Up to the current 2020 wave, HBS has been implemented by the MS, some EFTA countries and EU candidate countries on a 'gentlemen's agreement' meaning that HBS data have been transmitted to Eurostat on a voluntary basis. Each country kept the targets, the uses and the programming of its national HBS and, at the same time, it collaborated with Eurostat to compile a European-wide data set on household budgets with a five-year frequency and in accordance with commonly agreed methodological guidelines and recommendations.

However, not all countries fully aligned to the HBS guidelines. Consequently, Eurostat often received datafiles after the deadlines for submission, containing missing data, and with different file contents from countries. Datafiles were transmitted in a multitude of formats, which significantly hampered the process of validation, processing and preparation of indicators released in the Eurostat database, which was carried out using a tailored corporate tool based on the SAS software. Therefore, special ad hoc programmes and routines had to be written to read-in the data into Eurostat's IT (Information Technology) system.

In the first three waves of HBS implementation, countries delivered only aggregated tables. Since 2005, most countries (and since 2015 all countries) have delivered micro-data to allow Eurostat to perform an ex-post harmonisation and to answer specific requests of the users. Starting with the 2010 wave, Eurostat has released anonymised HBS micro-data files to the research community.

3. Impact of Regulation (EU) 2019/1700 on quality

In this section we introduce the new legal framework that will regulate HBS from the next wave in 2026. We describe the main regulation and the delegated and implementing acts and analyse their impact on relevance, accuracy, timeliness, accessibility and comparability of HBS data.

Starting with the 2026 wave, HBS will be implemented under Regulation (EU) 2019/1700 and related legislative acts. Regulation (EU) 2019/1700 covers seven domains of social statistics based on data at individual level collected from samples, and it specifies the content of the data, metadata information to be collected and the fundamental quality requirements as well as deadlines for transmission to be met.

Additional legislative acts, which concern all seven domains, have also been adopted pursuant to the framework regulation: e.g. Regulation (EU) 2019/2181 specifying technical characteristics as regards items common to several datasets; Regulation (EU) 2020/256 establishing a multiannual rolling planning; Regulation (EU) 2019/2180 specifying the detailed

arrangements and content for the quality reports. The full list of references can be found in the Appendix.

The new legal basis is expected to enhance the harmonization of HBS data among European countries and improve its cross-country comparability. Comparability over time should primarily be ensured in the future but it is expected that there will be some breaks in time series compared to previous waves, especially for the countries that will need to adapt their national methods to comply with the new legal requirements.

Furthermore, the new legal basis is expected to improve timeliness and punctuality: as already mentioned above, until the HBS 2020 wave, the data collection and data transmission to Eurostat have been done on a voluntary basis, and unfortunately, countries have not followed strictly the fixed deadlines for both data and metadata transmission. The new legal basis specifies the reference years, and data and metadata deadlines for each of the seven statistical domains covered by Regulation (EU) 2019/1700. In particular, under Commission Delegated Regulation (EU) 2020/256 establishing a multiannual rolling planning, countries shall conduct the data collection for the consumption domain in 2026 (see Annex I of the Regulation) , and according to Regulation (EU) 2019/1700, data shall be transmitted to Eurostat within 15 months of the end of the reference year (Annex V, (7)), and metadata within 3 months of the deadline for transmitting the data (Art. 13).

Additionally, the use of modern means of data collection and streamlining of validation procedures previewed for the HBS survey (to be described in the following sections of the paper) may contribute to additional improvements in timeliness.

However, the periodicity of HBS remains five years, which is a minimum requirement for updating the weights of the basket of goods and services used for the calculation of the Harmonised Indices of Consumer Prices – HICP given that a substantial number of countries do not plan to move to a yearly operation. HBS will therefore not be able to respond to immediate policy needs.

The new legal basis also provides general provisions on quality and its assurance, including the obligation for Eurostat to assess the quality and inform users about the findings. Several acts contain specific legal provisions regarding the content, structure and quality criteria for the national quality reports providing the metadata information, which are expected to improve data accuracy and precision.

In particular, Regulation (EU) 2019/1700, Annex II specifies the precision requirements for all data sets, including for the consumption domain, expressed in terms of standard errors for the

estimate of a particular indicator (for the consumption domain, the indicator is “Percentage of households whose expenditure on housing-related categories including water, electricity, gas and other fuels, are more than 50 % of the total expenditure”).

Furthermore, Regulation (EU) 2019/1700, Art. 13 specifies the quality requirements and metadata transmission deadlines to allow for the production of data that are comparable and compliant (e.g. countries are to provide metadata describing the methodology used, including the data sources and methods used, and how the technical specifications were achieved by reference to those laid down by the Regulation; information on compliance with the minimum requirements for the sampling frames used; information about the sub-populations that have not been reached by the data collection, etc.). Additionally, a specific implementing regulation specifies the detailed arrangements and content for the quality reports pursuant to Regulation (EU) 2019/1700, namely Regulation (EU) 2019/2180. The annex to the latter regulation sets out the quality criteria and statistical concepts in detail. They have been further detailed in Annex III of the HBS Implementing regulation, i.e. Regulation (EU) 2022/2094.

To further support Member States in improving the quality of the data sets, the new legal basis allows for feasibility and pilot studies to be launched (Regulation (EU) 2019/1700, Art. 14 and 16). They would cover data topics of comparability, developing innovative technologies and methods, modernising data collection and new users’ demands. For more details on supporting Member States in this regard see section on the innovation path: innovative tools and sources for HBS.

Referring to accessibility and clarity, the new legal basis sets an obligation to publish the aggregated data on the Commission (Eurostat) website and envisages granting access for scientific purposes to anonymised microdata. Eurostat will work on providing more statistics (see section on the impact of HBS modernisation and innovation path on quality for details) and related information in a more timely and user-friendly way, focusing on the electronic means of dissemination.

Last but not least, further harmonisation will be achieved with the six other EU social surveys by providing concrete common specifications regarding statistical concepts and definitions, statistical populations and observation units, the technical specifications of the data sets (regarding the content: description of variables, and the survey methodology: sample characteristics, precision requirements, or standards for data processing), the technical formats for the transmission of information, and characteristics of the sampling frames. These common requirements should improve the reliability of the data, and their coherence with e.g. the EU Statistics on Income and Living Conditions – EU-SILC and HICP.

To facilitate the implementation of the HBS survey under the new legal basis and to contribute to ensuring more harmonisation and better quality, Eurostat, in cooperation with the Member States, has elaborated a new package of supporting documents, including a methodological manual, data delivery guidelines, guidelines on data validation rules, indicators' manual and guidelines for national quality reporting. These documents will also help data users to better understand and use HBS data.

In conclusion, the full implementation of the body of legal provisions is expected to improve the quality of consumption expenditure data and lead to better compliance.

4. The impact of HBS modernisation and innovation path on quality

4.1 The modernisation path: a new production system

In this section we introduce the Eurostat project on the development of a new data production system for HBS. First we describe the elements: processes and tools like EDAMIS, STRUVAL, CONVAL, COOL and GSAST. Then we assess how the new system will impact the quality of HBS data through improvements in validation, processing and dissemination.

The current production system for HBS used by Eurostat consists of a SAS-based application in which all the validation and compilation of HBS microdata, and the production of output files is performed. The system has been used for HBS waves 2015 and 2020 with no major recent developments. It was not able to deal with all country-specific differences in data format and structure and therefore some country data needed to be pre-processed outside the system. In addition, all the production steps needed to be handled manually, which required substantial internal and external communication and was therefore time-consuming.

A redesign of the internal production system was launched to accommodate the requirements set in the new legal basis under which HBS will be conducted from 2026. The 'Modernisation of the HBS production' project will provide a new and robust mechanism for statistical production of the new HBS data submitted by the countries. The project aims to improve the quality, timeliness, cross-country comparability and relevance of the disseminated consumption expenditure indicators, taking into account the modern ways of data collection which are ever-more frequently used by countries. It will improve statistical production efficiency and increase the range and quality of statistical products.

This redesign will consist of introducing commonly used Eurostat tools for the following statistical processes: data transmission (from data providers to Eurostat), data validation, data compilation (including calculation of indicators) and production of various output data sets that will be used for dissemination purposes.

The transmission of HBS 2026 data and metadata from EU Member States and other countries will be arranged (as it currently is) via the Electronic Data Files Administration and Management Information System (EDAMIS), which is an electronic portal/interface and is referred to as the Single-Entry Point of Eurostat. A new automatic validation flow will be set up in EDAMIS, which means that each dataset is validated, and a validation report automatically sent back to the sender. The validation will be performed in two steps:

- structural validation (data structure definition - DSD based) in STRUVAL service.
- content validation (rule based) in CONVAL service.

The use of the automatic validation via EDAMIS, with a possibility to pre-validate the data before official transmission, should facilitate the validation process and make it more efficient compared to the manually managed validation performed until the wave 2020.

To access information on data transmission and on the data validation process, Eurostat domain managers can use the COOL (Corporate Orchestration and Overview Layer) application. This web application enables Eurostat users to monitor and manage several aspects of data exchange and validation and as such to facilitate the validation process and gain efficiency.

A new application for subsequent processing of the HBS microdata and production of data files for various data users will be developed. This SAS-based application (Generic SAS Tool - GSAST) will enable loading of data, including HBS 2026 microdata but also microdata and aggregated data from previous waves of HBS (1988-2020). It will also provide additional validation of the data, focusing on basic data description and relationships in the data to facilitate statistical analysis. Imputation of missing data and calculation of derived variables will be applied, which should improve the analytical power of the data.

The new tool will offer users an increased portfolio of indicators (for example median consumption expenditure or the proportion of households spending more than a certain percentage of their total consumption expenditure on specific consumption purpose) and will implement new breakdowns (e.g. tenure status, gender or disability status of the reference person). It will enable, where possible, the processing of historical microdata and aggregated data so that comparable time series can be produced. Increased parametrization features of the tool will provide data managers more flexibility and will reduce the need for updates of the system. Main outputs of the application will be datasets to be disseminated in Eurostat's statistical database (Eurobase) and user databases (UDB), which will be used, for example, to create anonymised microdata available to researchers. The application will produce various

reports to assess the quality of the original and output data and to facilitate the statistical analysis of the data. The system will also enable monitoring and assessing of the production processes.

4.2 The innovation path: innovative tools and sources for HBS

A third line of intervention, aimed to achieve higher data quality, is the project on innovative tools and sources for HBS. It was launched in 2017 and is supported by the work of a specific task force on innovative tools and sources for HBS created in the same year (Eurostat, 2023). The task force gathers national experts on innovation and modernisation issues. They have a mandate to work towards the modernisation of the HBS, providing recommendations to the working group on Income and living conditions and contributing to the European Statistical System (ESS) Innovation Agenda. Over the years, the task force has become a forum where Members States' experts can discuss innovation topics, present national projects' results, exchange best practices and experiences and share information.

In this framework, Eurostat has provided financial support to several European countries to improve the methodology and develop innovative and shareable solutions aiming at the modernisation of HBS data collection at national level. Among the financed projects, the most promising were the two projects led by Statistics Belgium: 'Software outreach, redefinition of flows and collect e-data through MOTUS (SOURCETM) and its follow-up project 'Establishing a cross-domain data collection platform for the ESS (CROESS)'; and the two projects '@HBS - An app-assisted approach for the HBS' and @HBS2, developed by two consortia led by Statistics Netherlands.

The diary is the core element of the survey. It collects details of the income and expenditures during a period stretching from 1-2 weeks in most countries to three months. The household budget surveys examine the household. Therefore, a household diary is to be kept for the household over that period. Individual diaries are also offered in many countries to allow the individual members of a household record their entries as they arise.

The diary is the most burdensome part of the survey from the respondents' perspective. Consequently, the innovative solutions - or tools' have been developed with the aim of reducing the response burden and reverse the declining trend in response rates seen in many European countries. They consist of web and smartphone applications covering both the questionnaire and the diary elements and include innovative (or smart) features such as the scanning and processing of shop receipts via optical character recognition - OCR (normally performed in the application's front-end). Instead of writing down all the expenses one by one on the diary, the respondent is asked to take a picture of the purchase receipt and the app will do the rest. In

the applications back end, machine learning algorithms would then classify the expenses according to COICOP categories based on the purchased products description.

As a result of development and testing efforts, fully fledged web and smartphone applications are currently available and some countries plan to use them in the 2026 data collection wave either as the only collection mode or in a multimode setting. Nevertheless, notwithstanding the availability of mature tools, many countries are still reluctant to adopt them, preferring a do-it-yourself approach. Understandably, there are still several challenges, such as shareability, trust, privacy preservation and usability, linked to any innovative solutions and not only to the available tools, which merit further research.

4.3 Towards (trusted) smart Household Budget Surveys

To tackle these new challenges, the focus of Eurostat projects has shifted towards the conceptualisation, proof of concept and implementation of (trusted) smart surveys (Ricciato and Wirthmann, 2019 and Ricciato et al., 2019). A smart survey involves a smart device (smartphone, tablet, activity tracker) to collect a combination of (inter)active data provided explicitly by the respondent with passive data collected in the background by the sensors. A trusted smart survey adds to the concept technological solutions aimed at increasing the degree of trustworthiness and strong protection of personal data based on privacy-preserving computation solutions.

Two of these ESSnet projects (the project ‘Trusted Smart Surveys’ and its follow-up ‘Smart Surveys Implementation’) build on the results of the innovative tools and sources project and have chosen HBS as a use case. The fact that HBS suffers from high respondent burden (and consequently low response rates), and recall and measurement issues, makes it a perfect candidate that can benefit from embedding smart features into its collection phase.

5. Final remarks

In this paper we discussed the strategy of Eurostat on HBS articulated around three distinct levels of intervention – legislative, process modernisation and survey innovation. The strategy is expected to have a significant positive impact on the quality of consumption expenditure data and indicators at national and EU level.

Firstly, at the legislative level, the new legal basis will ensure better availability of data and metadata (coverage with data for all countries) and improvements in quality for various aspects such as relevance, accuracy and reliability, comparability and coherence, timeliness, accessibility (legal requirements and new technical and methodological documents).

Secondly, the project on the development of a new data production system for HBS will modernise and streamline the validation, processing and dissemination processes in Eurostat, which will contribute to quality improvements (accuracy, comparability, timeliness among others) and efficiency gains in the production of HBS statistics and relating metadata.

Thirdly, the innovation on data collection tools and recent achievements on the use of alternative data sources are expected to have a significant impact on data quality in terms of reduction of recall and underreporting errors, real-time data checks and pre-validation (performed in-app or in the back-office) and reduced post-fieldwork data processing time. Such developments would result in improved overall data quality in terms of timeliness and accuracy, while maintaining data relevance and preserving privacy and confidentiality.

References

Eurostat (2023). Innovative Tools and Sources for the Household Budget Survey. *Report to the working group Income and living conditions - Household budget survey*. Available on request.

Minnen, J., E. Nagel, & K. Sabbe (2020). SOURCE™: Software Outreach and Redefinition to Collect E-data Through MOTUS. Towards a Modular Online Time Use Survey. *Final report.* Available online (link).

Ruppert, E., F. Grommé, F. Ustek-Spilda, & B. Cakici (2018). Citizen Data and Trust in Official Statistics. *Economie et Statistique / Economics and Statistics*, 505-506, 171–184. https://doi.org/10.24187/ecostat.2018.505d.1971

Ricciato, F. & A. Wirthmann (2019). Trusted Smart Statistics: how new data will change official statistics.

Ricciato F., A. Wirthmann, K. Giannakouris, F. Reis, and M. Skaliotis (2019). Trusted smart statistics: Motivations and principles. *Statistical Journal of the IAOS 35*, 589–603, DOI 10.3233/SJI-190584

Sabbe K., J. Olsen, E. Bitzer, E. Nagel, B. Lenuweit, D. Knapp, C. März, A. Sommer, P. Brecht, J. Minnen, P. Beyens, and E. Cortebeeck (2022). CROESS Establishing a Cross-Domain Data Collection Platform for the ESS. *Final report*. Available online (link).

Schouten B., T. Oerlemans, N. de Wolf, B. Vrabič Kek, S. van Hoek, G. van den Heuvel, J. Bos, É. Ménesi, M. Zanatyné, M. Zgonec, S. Theunissen, J. Akkermans, E. Rodenburg, H. Attila, T. Parikka, Č. Grahonja, N. Kebe, & A. Elevelt (2022). @HBS2: An app-assisted approach for the Household Budget Survey. *Final report*. Available online (link).

Appendix - List of HBS 2026 legislation

- Regulation (EU) 2019/1700 of the European Parliament and of the Council of 10 October 2019 establishing a common framework for European statistics relating to persons and households, based on data at individual level collected from samples, amending Regulations (EC) No 808/2004, (EC) No 452/2008 and (EC) No 1338/2008 of the European Parliament and of the Council, and repealing Regulation (EC) No 1177/2003 of the European Parliament and of the Council and Council Regulation (EC) No 577/98: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R1700&from=EN
- Commission Delegated Regulation (EU) 2020/256 of 16 December 2019 supplementing Regulation (EU) 2019/1700 of the European Parliament and of the Council by establishing a multiannual rolling planning: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32020R0256&from=EN

- Commission Implementing Regulation (EU) 2019/2181 of 16 December 2019 specifying technical characteristics as regards items common to several datasets pursuant to Regulation (EU) 2019/1700 of the European Parliament and of the Council: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R2181&from=EN
- Commission Implementing Regulation (EU) 2019/2180 of 16 December 2019 specifying the detailed arrangements and content for the quality reports pursuant to Regulation (EU) 2019/1700 of the European Parliament and of the Council. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R2180&from=EN
- Commission Delegated Regulation (EU) 2023/126 of 21 October 2022 supplementing Regulation (EU) 2019/1700 of the European Parliament and of the Council by specifying the number and the title of the variables for the consumption domain: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32023R0126&from=EN
- Commission Implementing Regulation (EU) 2022/2094 of 28 October 2022 specifying the technical items of data sets, establishing the technical formats for transmission of information and specifying the detailed arrangements and content of the quality reports on the organisation of a sample survey in the consumption domain pursuant to regulation (EU) 2019/1700 of the European Parliament and of the Council: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R2094&qid=1676543640717&from=EN
- Commission Implementing Regulation Decision (EU) 2023/2711 of 5 December 2023 granting derogations to a Member State from Regulation (EU) 2019/1700 of the European Parliament and of the Council, as regards the delivery of data on certain variables in statistics relating to persons and households: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202302711