Online household diary in Hungarian Household Budget Survey. Magic kit or mistake

Éva Ménesi (Central Statistical Office, Hungary)

eva.menesi@ksh.hu

Abstract

Household Budget Survey is said to be a good, old fashioned tool of statisticians to collect information on expenditures and spending structures of private households. Keeping diary in paper form in the 21st century seems to be outdated, time consuming and putting heavy burden on respondents as well. This presentation and paper tries to give a picture of the efforts taken by the Hungarian Central Statistical Office to modernize the data collection using online e-Dairy.

HU-HBS is an annual survey. Sample size is about 8000 households. Data collection period is two weeks and the sample is distributed equally during the year. Paper diary (PAPI) was the standard data collection method but since 2015 e-Diary CAWI) has been developed and offered as an option for respondents. The online diary uses a special IT infrastructure developed by a contractor for the purpose. The tool offers predictive data entry, user friendly structure and tailor made modular system. To collect information on the household structure a personal interview is carried out by interviewers using laptops (CAPI). The collected data from all data collection methods (CAWI, PAPI, CAPI) are merged into a single dataset.

Incentive is offered for household to switch to online version but according to the current situation the share of online diary is very low. To make it more favorable for respondent is our challenge. It is investigated how aggregated and useful information could be provided for household filling the e-Diary and make PAPI respondents more eager to switch to CAWI.

Keywords: Household Budget Survey, modernize, web-diary, CAWI

Online household diary in Hungarian Household Budget Survey

Magic kit or mistake
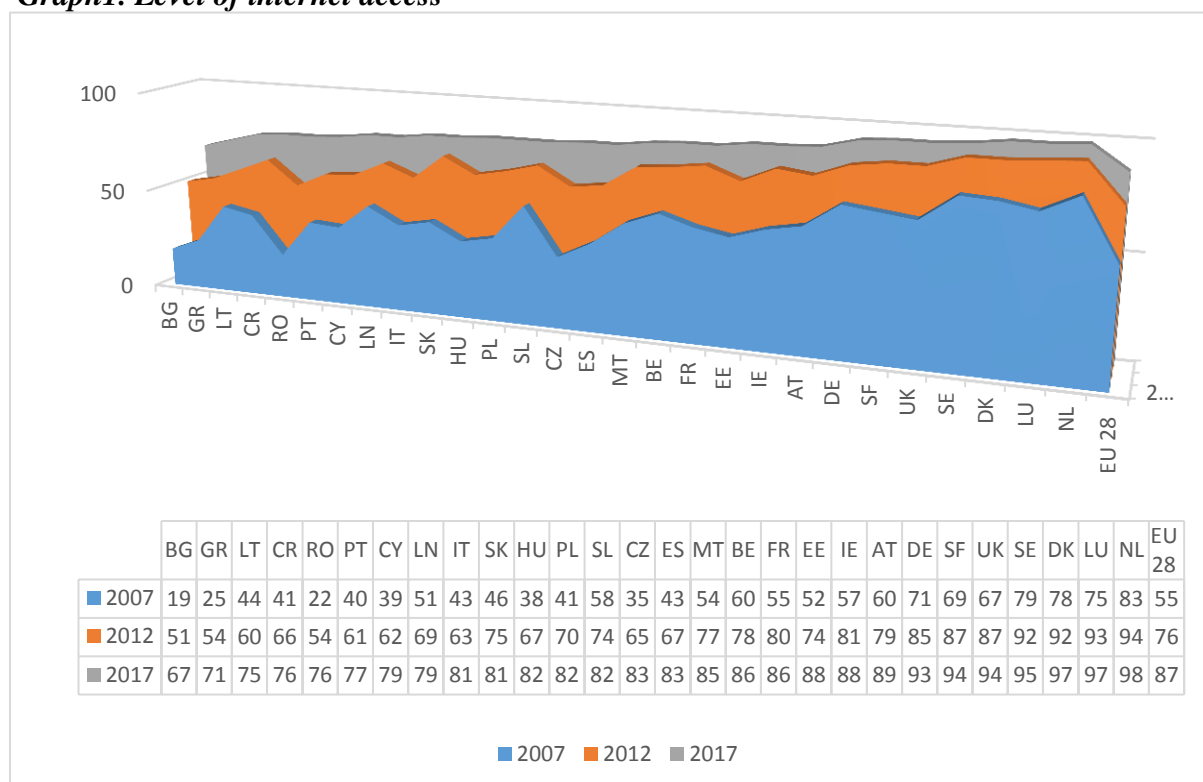
Éva Ménesi
senior statistician

Department of Quality of Life Statistics
Hungarian Central Statistical Office (HCSO)
HU-1024 Budapest Keleti Károly u. 5-7.
eva.menesi@ksh.hu

1. Introduction

The Digital Revolution, (or as it often called as the Third Industrial Revolution**)** started in the 1960-1970s with the adoption and proliferation of digital computers and digital record keeping that continues to the present day. This phrase express those fundamental changes which effects our current life and how computer technology became a fundamental building block of everyday life. Central to this revolution is the mass production and widespread use computers, digital cellular phone, and the Internet by the beginning of XXI. Century. The Digital revolution is said to be the beginning of the Information Age.

The notion of the digital revolution is part of the Schumpeterian theory of socio-economic evolution, which consists of an incessant process of creative destruction that modernizes the modus operandi of society as a whole, including its economic, social, cultural, and political organization.

Graph1. Level of internet access

	BG	GR	LT	CR	RO	PT	CY	LN	IT	SK	HU	PL	SL	CZ	ES	MT	BE	FR	EE	IE	AT	DE	SF	UK	SE	DK	LU	NL	EU 28
2007	19	25	44	41	22	40	39	51	43	46	38	41	58	35	43	54	60	55	52	57	60	71	69	67	79	78	75	83	55
2012	51	54	60	66	54	61	62	69	63	75	67	70	74	65	67	77	78	80	74	81	79	85	87	87	92	92	93	94	76
2017	67	71	75	76	76	77	79	79	81	81	82	82	82	83	83	85	86	86	88	88	89	93	94	94	95	97	97	98	87

2007   2012   2017

Percentage of households who have internet access at home. All forms of internet use are included. The population considered is aged 16 to 74.
Source: Eurostat ICT survey

In this context official statistics need to update their portfolio of data collection tools to reflect the fundamental changes in the society. Existing and new data collection methods have to be handled and even matched, cost effective and environment friendly approaches have to be developed and maintained. This paper will describe the effort has taken by the Hungarian Central Statistical Office (HCSO) to improve and modernize the data collection of Household Budget Survey (HBS) since 2015.

2. Household Budget Survey in HCSO

HBS has got very long roots in the Hungarian statistical data collection system. The first survey was carried out in 1949. After about 70 years it aims basically the same purpose of collecting information about the expenditures and income of Hungarian private households. During this life-long period the survey had to be continuously adapted to the changing of social and economic situation of the society and always bear in mind the possibility of producing reliable and comparable data for internal and external users.

From the beginning of the nineties HBS data are used for the following purposes. It is a source of information on structure and level of household consumption expenditures. It provides data for estimation of GDP household consumption side. It gives the weights for the Consumer Price Index (CPI) calculation.
From a broader point of view, from the EU perspective besides the above mentioned aims the Commission uses HBS data mainly for policy purposes like consumer protection, regional policy or Purchasing Power Standard (PPS) comparisons.

The survey has already gone through many modernization processes regarding the data collection framework and questionnaire design, sampling, etc.
After the political transition in 1989 the new requirements of the market economy had to be reflected in HBS. The response-rate was getting lower in certain social groups which required more sophisticated imputation techniques. We adopted the EU regulations step by step. At first our attention was focused on the sampling methods problems and imputation techniques. It involved a detailed standard error calculation and several publications on the sample design.
By 2003 we adopted all the recommendations regarding the classification on consumption such as E-COICOP and income classification which were the same as recommended for EU-SILC survey. We report the expenditure data on COICOP HBS 5 digits level while for internal purposes we have the deeper 6 digit level breakdowns as well.

The data collection period was shortened from the original 1 month to 2 weeks; the household diary became structured; the connecting retrospective interview on greater purchases was modularized. So at the current stage HBS-HU complies with the EU recommendations but still keeping some national features. E.g. For food we collect data on the quantities of consumed food items and report them in the corresponding measurements units like kilogram, liter or pieces. It provides extended possibilities on analysis and food safety studies, etc. We also collects data on own produced goods (vegetables, fruits, firewood produced in the garden/farm of the household), although its share from the total consumption is declining rapidly.

This paper focuses on the most recent facelift of the survey namely the introduction of online web diary.

As a state of play in the beginning of 2010s all of the social survey in HCSO were still using paper-and pencil interviewing. Labour Force Survey (LFS) introduced CAPI interviewing in its second wave. It was the capacity building period when a necessary IT devices (laptops, servers) were purchased and the IT infrastructure was developed. The first online questionnaire was used for the Census 2011. We were delighted because

about 20 % of the respondents has chosen the CAWI version for the responding. This favorable reception has opened the doors for web questionnaires in other social surveys.

By 2016 paper questionnaires were not used anymore. Laptops and tablets were used for CAPI interviewing and CAWI version were developed for all social surveys.

Table1. Data collection modes in social survey at HCSO in 2011-2012

	PAPI	CAPI	CATI	CAWI
LFS 1st wave	X	X		
LFS 2nd +later waves		X	X	
SILC	X			
HBS	X			
ICT	X			
Census	X			X

Table2. Data collection modes in social survey at HCSO in 2015-2016

	PAPI	CAPI	CATI	CAWI
LFS 1st wave		X		
LFS 2nd +later waves			X	X
SILC		X		X
HBS	X	X		X
ICT		X		X
Microcensus		X		X

3. Introducing online web-diary in HBS

In 2015 a special IT infrastructure was built for the purpose of HCSO to carry out data collections. The framework was developed by a contractor and maintained by our IT experts.
The system contains a questionnaire design feature, which can be operated by statisticians. The system is connected to a monitoring database which contains the selected sample and able to follow the evolution of the data collection process. It shows whether each selected household has already been interviewed, or has they already started the online diary. This IT system also responsible for the data sending to regional statistical centers and then to the central office.

Fundamentally the IT framework was developed for CAPI data collection method and then was improved to serve for CAWI purposes.

3.1. Paper diary

The current HBS-HU is a data collection which include a 2 week diary keeping period. The households are asked to record all their consumption expenditures during the diary keeping period, keeping record of their food purchases in quantity as well. The paper diary is well structured.
The first chapter is dedicated to *Important expenditures during the 2 weeks of the data collection*. E.g. Eating out of Home (school/workplace, home delivery of lunch,

restaurant, coffee, soft drinks), Wine-Beer, Clothing, Cost of transportation, Medical expenses, Hairdressing, entertainment, gardening. While the second chapter – *Daily expenditures* contains 14 pages with some prefilled items (bread, cheese, meat, fish, fruit, and vegetables) and free entry rows for other items. The third part a supplementary page with list of items for *Monthly expenditures* (housing, transport, telecommunication, education, insurances, savings, and mortgage / loan repayments).

The paper diary is coded according to COICOP classification by the interviewers after the data collection period.
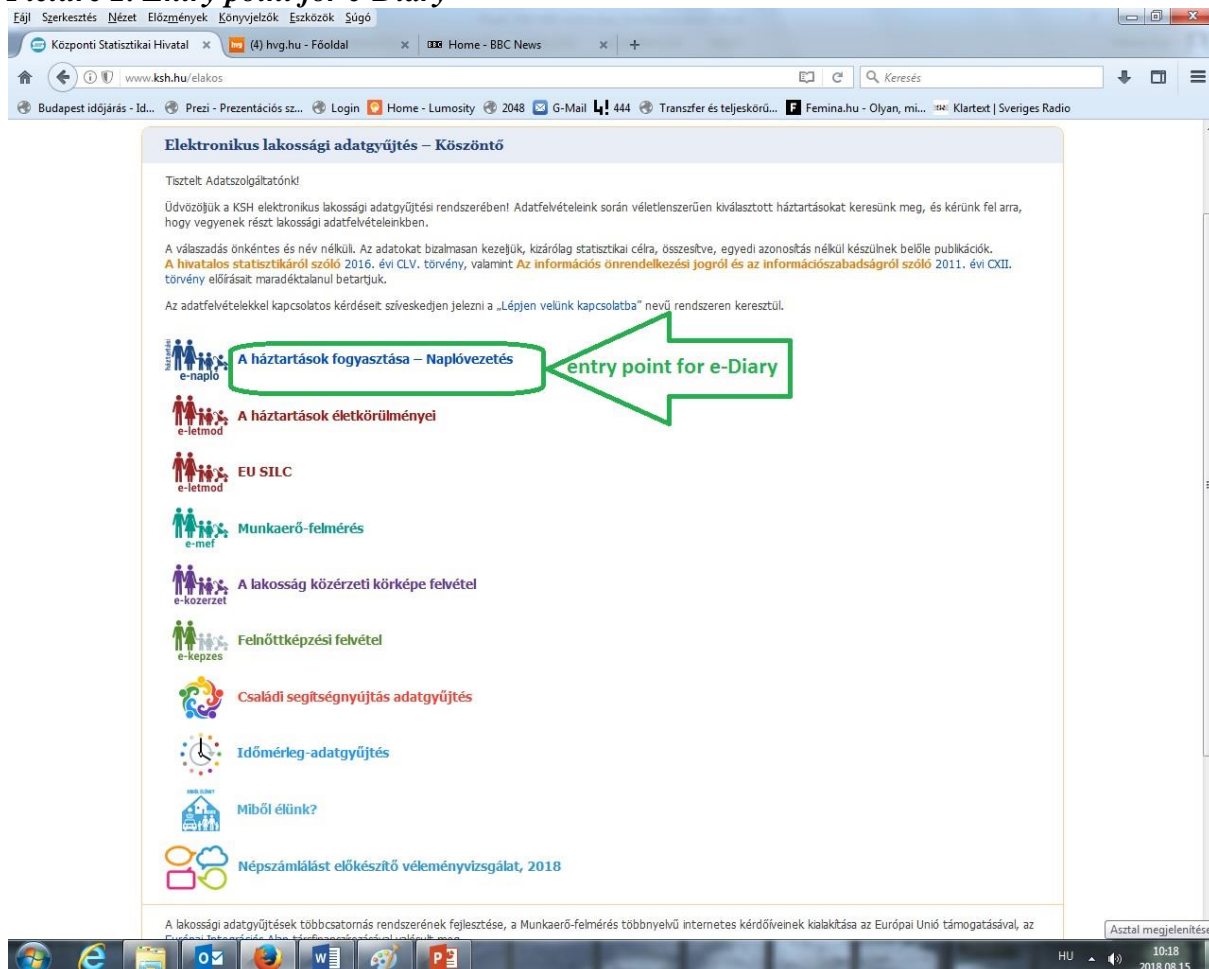
3.2. e-Diary

The paper diary was the starting point for the construction of web version. Our aim was to come up with a user friendly tool which contains built in check routines, messages, predictive data entry and online coding and follows the structure of the paper version.
The framework programming was done by a contractor in 2015. Any modification, changes are done by our colleagues at HCSO since that time.

The e-Diary can be reached by the respondent on the main page of HCSO thought a common entry point.

Picture 1. e - Questionnaires entry point for all social surveys

Picture 2. Entry point for e-Diary

The e-Diary start with a welcome page where the main purpose of the survey is described, legal references, contact information are also available here. A toll-free phone number and an e-mail address are also provided for the respondents in case of any questions.

The web version has 3 main chapters. It starts with the block of monthly expenses. The second one is dedicated to daily expenses, while the third one covers the consumption of own produced goods.

The items in each chapters are listed on the basis of the frequencies occurred in the paper diary, so the most frequent items are listed at the top while the less frequent items listed at the bottom. E.g. electricity and gas costs are listed at the top, since most of the households have these types of expenditures in a month, while private sewage cleaning is listed at the bottom, since the public sewage cleaning is the standard in most of the households.

All categories are in accordance with COICOP classification. The respondents can fill the diary with simply clicking on the item on the left column, or she/he can enter an item on her/his own. The rows offers predictive data entry possibility, so when the respondents starts a word with some letter the program offers her/him items starting with the same letters, and she/he can choose from the list. This predictive entry speed up the filling process.

All the chapters have built in checks e.g. upper and lower limits for costs and quantities.
Detailed pictures of the e-Diary can be shown during a presentation to illustrate the above mentioned elements.

The monitoring system allow us to check the status of the e-Diaries. Whether the respondent has already started the filling process, or stopped for more than 3 days, or finished successfully.
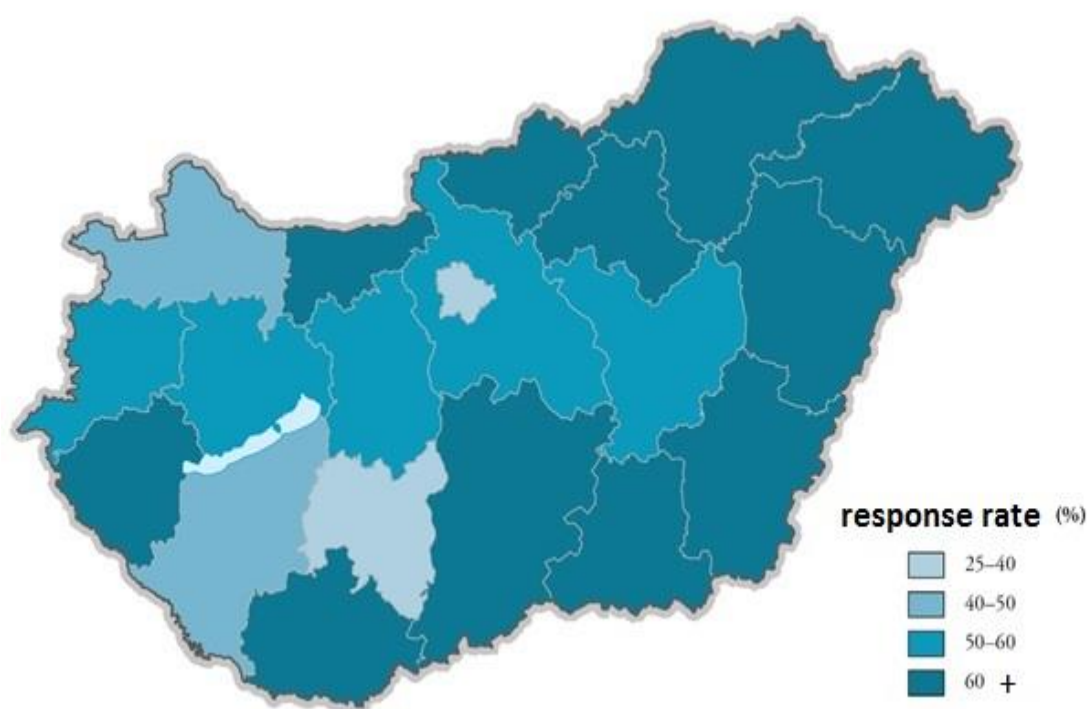
After a successful e-Diary completion the interviewer goes back to the households and ask a short questionnaire about the structure and characteristics of the household and its members. At the same time the incentive is given to the household.

Regarding incentives there is no difference between the data collection modes. Due to taxation reasons HCSO is unable to pay for the respondents, so in kind gifts are given. E. g. pack of coffee and a stationary product (pen, pen drive) or shopping bag, mug with the logo of HCSO.

Interviewers are paid by the number of successful interviews regardless of the data collection mode.

Data provided by the e-Diary (CAWI) and the household characteristics questionnaire (CAPI) are merged at the regional statistical centers. The process is the same for paper diary. The coding and data capturing is done at the regional centers and the dataset is merged with CAPI file of household characteristics. In the center office the files of both data collection modes are merged into one single data set in SAS.

Graph 2. Overall response rate for HBS, 2015

The overall response rate for HBS is 53 % on country level, while urban areas (especially the capital city Budapest) and western, more well off parts of the country show much lower response rates.

Table3. Number of diaries by mode of data collection in HBS

	PAPI	CAWI
2015	9172	122
2016	9941	121
2017	9478	129

Share of e-Diary is really very small compared to the standard paper diary from its introduction till the present year.

Table4. Number and share of CAWI in HBS and SILC

CAWI	HBS-diary	% of the total	SILC	% of the total
2015	122	1.3	265	1.7
2016	121	1.2	958	5.8
2017	129	1.3	1106	7.1

Comparing the achievements in CAWI in HBS and SILC the result show extensive difference either in the number of completed questionnaires (diaries) or in the share of CAWI from achieved sample. HBS e-Diary is underperforming compared to SILC.

4. Conclusions

The e-Diary as a new online data collection tool which was introduced to the Hungarian data collection system in 2015.

From the data provider perspective the tool offers many user friendly features – e.g. predictive data entry, error and completeness messages.

From the statistical institution perspective it considerably speed up the data collection and increase the quality of COICOP coding process. It shortens the evaluation period, and improve the quality and consistency of the recorded data. The inclusion of the online monitoring system provides up-to-date information on the evolution of the achieved sample over time. After the capacity building period it is cost effective. Any modification to the standard e-Diary can be constructed by the statistician.

Besides those so many value gains why this new and modern tool is still underperforming in the sense of share of successful interview compared to the paper version?

The online monitoring system allow us to follow the evolution of the e-Diary performance since the sample selection till the achieved sample size. This feature was used to analyze how many households have agreed to answer by the e-Diary and had started the filling process but have not finished it (incomplete Diary).

Table 5. Households agreed to fill the e-Diary by completion level in 2017.

Diary completed	129	90,8%
Diary opened, no record at all	9	6,3%
Diary filled, too few records	4	2,8%
Total	142	100,0%

The vast majority of household who agreed to fill the e-Diary had completed it and their data were used in the dataset. There were very few households who filled very limited number of items and we were unable to use their answer in the datasets. While about 6 % of the households opened the tool but have not done anything. In our understanding those households used CAWI option to camouflage their unwillingness to cooperate and non-response to the survey at all.

So we think the tool seems to be user friendly and the attitude of respondents to CAWI should be encouraged.

In SILC survey a lottery is offered for CAWI respondents and 3 tablet can be won by 3 lucky persons.

Similar approach will be followed here to offer a lottery to win 1 tablet for the households filling e-Diary to promote its use in 2019. Meanwhile greater efforts should be put on interviewers to persuade households to switch to online version instead of paper diary. Some kind of extra compensation premium should be offered to the most successful interviewers on this activity.

From IT perspective we considering to put some efforts on a development of a smartphone application of the e-Diary in the near future.

5. References

https://en.wikipedia.org/wiki/Digital_Revolution

Carlota Perez, Leo Johnson, Art Kleiner. *Are We on the Verge of a Golden Age?*. August 28, 2017.

Schumpeter, J.A. *The theory of economic development : an inquiry into profits, capital, credit, interest, and the business cycle* translated from the German by Redvers Opie (1961) New York: OUP

Schumpeter, J.A. *An Introduction to Economics with Emphasis on Innovation*, Pol, E Carroll,P, 2006

UNECE Canberra Group Handbook on Household Income Statistics,2nd ed. New York, Geneva, 2011- Chapter3. 3.2. Source of household income statitics

Level of internet acess – Eurostat static tables:
http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tin00134&plugin=1