



UNECE Expert Meeting on Statistical Data Collection 12 – 14 June 2023

Paradata and Data Collection Techniques

Feasibility of the WEB data collection for the Slovenian Labour Force Survey

Prepared by: Hana Vratanar, Tina Osvald Zaletelj, Eva Belak, Katja Rutar, Marko Sluga

Hana Vratanar, Statistical Office of the Republic of Slovenia, hana.vratanar@gov.si
Tina Osvald Zaletelj, Statistical Office of the Republic of Slovenia, tina.osvald-zaletelj@gov.si
Eva Belak, Statistical Office of the Republic of Slovenia, eva.belak@gov.si
Katja Rutar, Statistical Office of the Republic of Slovenia, katja.rutar@gov.si
Marko Sluga, Statistical Office of the Republic of Slovenia, marko.sluga@gov.si

The main goal of the LFS WEB pilot survey was to prepare an appropriate questionnaire for the WEB and to test the WEB data collection method for the LFS. The pilot focused on the preparation and testing of the best possible solution for WEB data collection at the household level.

Our contribution analyses WEB pilot LFS data and paradata collected for the 5th Wave 2022. Our aim was to answer the following questions:

- On which devices respondents started to fill out the questionnaire and on which devices it was completed?
- How many respondents entered the questionnaire and left it before the end (drop-off rate)? Does the household size/type have any impact on the drop-off rate?
- Are the descriptions given by the respondents regarding occupation, economic activity and education detailed enough for subsequent ISCO, NACE and ISCED coding at the required level?
- Is it feasible to incorporate code lists for occupation, economic activities and education (if any) into the web questionnaire?
- What is the share of the proxy answers for the key questions and what is the share of “don’t know” answers by proxy respondents?

All the analysis will be checked by socio-demographic characteristics of the selected person and by some characteristics of the household to which the selected person belongs. At the end of our presentation, we will present our plans and possible scenarios of the WEB data collection in the regular LFS production.

Disclaimer

The paper represents the personal views of the authors and need not necessarily coincide with those of the Statistical Office of Slovenia.

I. Background

1. At SURS, we have implemented several surveys for individuals for the WEB survey mode. Labour Force Survey is household panel survey and we do not have experience with household panel WEB surveys at SURS.
2. LFS is the largest continuous household survey in Slovenia. In the first wave, respondents are interviewed using CAPI mode (face-to-face), while in the subsequent three waves they are contacted via telephone (CATI).
3. Participation in the Slovenian LFS is voluntary. Therefore, different approaches need to be implemented to tackle the decreasing response rates. WEB questionnaire, in this case, offers additional opportunity to address respondents who prefer online questionnaires and who otherwise would not have responded to the survey.
4. Main goals of the pilot survey were to prepare an appropriate LFS questionnaire for the WEB and to test the WEB data collection method for the LFS. The focus was primarily on the preparation and testing of the best possible solution for WEB data collection at the household level for the subsequent waves of the LFS data collection.

II. Questionnaire design for WEB LFS

5. In the case of LFS WEB survey questionnaire, we are supposed to collect data from all household members. In the introduction module of the questionnaire we had to introduce respondent-friendly questions to check the composition of the household, collected in the previous wave, and add new household members (newly born household members and persons that moved in the household) and remove household members that have deceased or moved out of the household.
6. Preparing questions for the repeated data collection. We used dependent interviewing for:
 - data on household members and their relationships,
 - person number, name, surname, gender, date of birth, month of birth, year of birth,
 - economic activity for main work (name of the business/firm, kind of products or services supplied by the firm/business),
 - occupation for main job (job title and job description),
 - economic activity for last work (name of the business, types of services that are produced),
 - occupation for last job (job title and job description),
 - highest education level attained, year in which they attained the highest level of education, field of education (4-digit code), indicator of whether the person is working or not.
7. Thus, the respondent checked some of the data in the questionnaire and in the case when the answers have changed from the previous wave, subsequent questions were asked.
8. Implementation of the 'no proxy' and 'proxy' questions. The format of the questions in the questionnaire is adapted depending on whether the person is answering for herself/himself (direct interview; questions and answers are presented in the first person singular) or whether another person is answering for a person in the household (proxy interview). In case of a proxy respondent, questions and answers are presented in the third person singular. Due to these adjustments, a lot of extra work was done.
9. Preparation of the introductory module for the household and questions on persons in the household. In the introductory module (ID Module), we had to check whether selected person, who defines the household, still lives on the address where we have sent the advance letter. If so, we then list household members from the previous wave, add new members, and delete members that are no longer in the household. There were no problems detected with the introductory module of the questionnaire.
10. Implementation of the coding lists. The Standard Classification of Occupations (ISCO-08), Standard Classification of Activities (NACE) and Standard Classification of Education, Fields of Education and Training (ISCED) were not implemented in the WEB questionnaire.

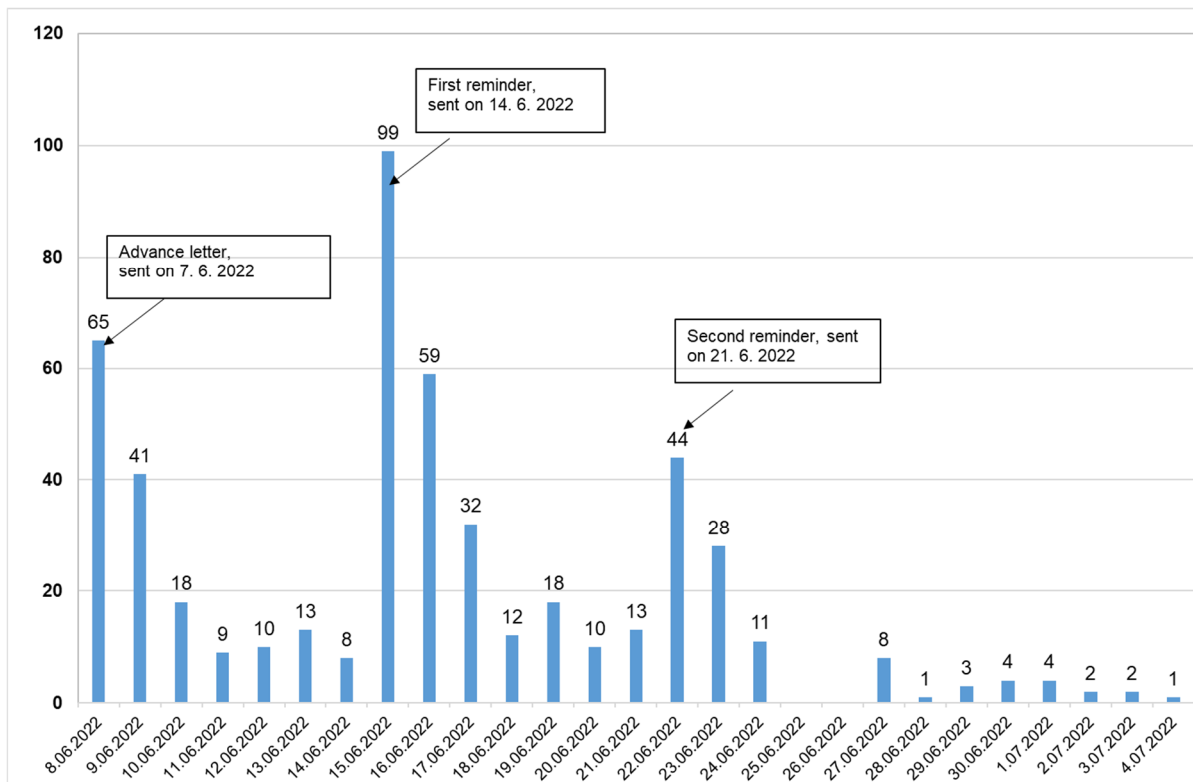
The code lists are too exhaustive to be included in the questionnaire. There are technical reasons as well as the usability of the code list in the questionnaire. Before implementing such a list in the questionnaire, separate experiment regarding the usability of such lists should be carried out.

11. In the questions, where the information regarding occupation, activity of the business and education is required, we have included examples in order to obtain description, detailed enough for subsequent coding.

III. Communication strategy and effectiveness of reminders

12. Analysis of the effectiveness of the advance letter and both reminders. The graph (Figure 1) represents the time when respondents answered the WEB survey. As we can see, the number of completed questionnaires was the highest around the time of sending the letters (advance letter, first and second reminder). Similarly as in other WEB surveys at SURS, the first reminder has the highest impact, which we sent 7 days after the advance letter. The first reminder sent on June 14, 2022, when 99 questionnaires were completed. According to the results, the strategy of sending the advance letter and both reminders is important.

Figure 1 Contact strategy – effects of advance letter and both reminders



IV. Response analysis

13. Characteristic of the respondents. We got responses from 48% of households in the sample. They reported 1394 household members, half men, half women, with average household size of 2.8 members. Their activity statuses were not calculated yet. When we look at households with unemployed members (46) in 5th wave, 52% of them responded in pilot WEB data collection while 48% of households with no unemployed persons responded in the pilot.

Table 1 How many persons completed the questionnaire? Sample composition of the initial sample in the 1st Wave, 5th Wave, which was WEB pilot sample and WEB Pilot response

Age group of the selected person	Gender	1 st Wave	%	5 th Wave WEB Pilot sample	%	WEB Pilot Response	%
18-24	M	347	4.7%	23	2.2%	14	2.8%
	F	333	4.5%	32	3.1%	14	2.8%
25-34	M	692	9.4%	55	5.3%	22	4.4%
	F	539	7.4%	48	4.6%	19	3.8%
35-44	M	859	11.7%	96	9.2%	51	10.3%
	F	717	9.8%	90	8.6%	50	10.1%
45-54	M	853	11.6%	97	9.3%	56	11.3%
	F	736	10.0%	132	12.7%	76	15.4%
55-64	M	760	10.4%	127	12.2%	54	10.9%
	F	786	10.7%	162	15.5%	78	15.8%
65-74	M	175	2.4%	44	4.2%	24	4.8%
	F	230	3.1%	63	6.0%	20	4.0%
75 or more	M	104	1.4%	30	2.9%	8	1.6%
	F	195	2.7%	43	4.1%	9	1.8%
Total		7326	100.0%	1042	100.0%	502	100.0%

14. Non-response analysis, information and response from the field. As stated above, we kept record of those persons, who informed us that they do not want to participate due to various reasons – there were 98 of them. Average age of persons who inform us that they do not want to participate was 67 years. 29% of them were men and 71% were women.
15. Reasons for 'not wanting to participate' were as followed: (i) refusal; (ii) absence; (iii) incapability; (iv) respondent does not have a computer/internet; (v) respondent does not know how to use a computer; (vi) unknown and (vii) moving to another address.
16. The most common reason for not participating (for those who contacted us) in the WEB survey was because the respondents did not have a computer or access to internet (72%) – the average age of these persons was 70 years. The second most common reason was hard refusal (13%). They commented that they did not have time to participate or that they had already participated several times and do not want to anymore.

V. Data collection results

17. Remarks on the WEB questionnaire at the end of the questionnaire. At the end of the questionnaire, we added two additional questions for respondents to answer. One question was about how they would prefer to answer such questionnaire in the future and in the other question, we asked respondents to describe their experience in filling out the WEB questionnaire and to suggest how we could improve the experience in the future.
18. The wording of the first question was: »You have gained experience by completing a WEB questionnaire. How would you prefer to answer such a questionnaire? « The results show that majority would prefer to answer such questionnaire on the WEB (72%). Some respondents would rather answer such survey over the telephone (10%) since they think it took them considerable less time and there is an interviewer on the other side who can additionally explain some questions if necessary. About 17% of the respondents do not care how the data is collected.

Table 2 Preferred mode of interviewing

	n	%
On the WEB	341	71.9
On the telephone	48	10.1
In-person (CAPI)	3	0.63
It doesn't matter how	82	17.3
Total	474	

19. The wording of the second question was “If you would like, please describe your experience filling out the WEB questionnaire and suggest how we could improve the experience in the future.” We grouped the answers into five groups as showed in the table (Table 3).

Table 3 Experiences with filling out the WEB questionnaire

	n	%
Respondents who were satisfied with the WEB questionnaire and have nothing to add	39	35.8
Respondents who filled out the WEB questionnaire but refused to cooperate in our survey in the future	22	20.2
Respondents who expressed their dissatisfaction with the length of the WEB questionnaire	9	8.3
Respondents who had problems filling out WEB questionnaire due to unclear questions or other dilemmas.	28	25.7
Respondents who provided technical suggestions for improving the WEB questionnaire	11	10.1
Total	109	

20. Among those who answered the open question there was 36% of those who were satisfied with the WEB questionnaire, 26 % had problems filling-out the questionnaire, 10 % had technical suggestions for improving the WEB questionnaire. Some of the suggestions were about adding a progress bar (showing percentage of completed survey) and having more ‘controls’ in the WEB questionnaire.

21. Many comments from the respondents were that the questionnaire is too long and too comprehensive. Some of them thought that there were too many similar questions – for example, regarding absence. Furthermore, some even thought that some questions are not relevant to their activity status.

22. Contact data and information regarding the completion of the questionnaire. LFS WEB questionnaire is the first WEB questionnaire at SURS where we collect the data about all household members. This was challenging since we had to collect data for all household members aged 15 years or more. Ideally, each person in the household would answer the questionnaire for himself. However, in practice most of the questionnaires were filled out by one person (82%), 13% by two and only smaller proportion by three or more persons.

Table 4 How many persons completed the questionnaire?

	n	%
Not completed	6	1.2
One person	424	82.2
Two persons	67	13
Three+ persons	19	3.7
Total	516	

23. At the end of the questionnaire, we asked for contact data for selected person and those household members that are aged 15 years or more and have answered the questionnaire by themselves. We asked for e-mail address and mobile and fixed telephone number.
- We informed the respondents that contact data is collected only for the purpose of this survey.
 - We have obtained e-mail addresses for 73% and mobile numbers for 71% of selected persons, while for other household members aged 15 years or more, who have filled out the questionnaire by themselves, we obtained e-mails for 24% of them and mobile numbers for 23% of them.

Table 5 Contact data obtained at the end of the questionnaire

	n	e-mail	%	Mobile phone	%
Self-report	610	468	77%	447	73 %
Other hh. members. aged 15+	600	27	5%	28	5%
Total	1210	495	41%	475	39%

24. Time to complete the questionnaire increases with household size. The presence of an interviewer considerably shortens the interviewing time.

Table 6 Time to complete the questionnaire by household size

	CATI * (4 th wave)	WEB PILOT (5 th wave)
Household size (members aged 15+)	Mean (in minutes)	Mean (in minutes)
One member	3.9	10.0
Two members 15+	5.6	14.0
Three members 15+	8.0	19.7
Four or more members 15+	9.7	23.6
Total	6.4	16.0

**The same households were considered.*

Table 7 On which device questionnaire was completed?

	Entered the questionnaire		Completed the questionnaire		Drop-off	
	n	%	n	%	n	%
Desktop/laptop			329	69.0		
Smartphone			137	28.7		
Tablet			11	2.3		
Total	516	100%	477	100%	39	7.5%

25. One of the indicators of the clarity and complexity of the questionnaire is also the share of item non-response and “don’t know” answers. Since WEB questionnaire does not provide interviewers assistance, non-response to some questions were significantly higher in WEB LFS pilot than in the regular LFS implementation (CATI). There was also difference in the share of “don’t know” answers in case of self-report and in case of proxy. All questions of the questionnaire were obligatory.
26. To questions, from which labour status was derived (employed, unemployed, inactive) “don’t know” answer was not allowed. Therefore, key questions for the analysis were selected from those who proved to be more complex, e.g. questions with open text fields on

economic activities and occupations, questions on the number of hours usually or actually worked and field of education attained.

Table 8 Percentage of DON'T KNOW answers for some key questions in case of self-report and in case of proxy

Key questions	DON'T KNOW Self-report		DON'T KNOW Proxy	
	n	%	n	%
Hours usually worked in the main job	15	12.7	46	30.5
Hours actually worked in the main job	31	8.4	73	22.7
Number of hours that person would wish to work	42	11.0	129	40.0
Field of education	/	0.0	1	2.9

27. Results of the LFS WEB pilot showed interesting “anomaly” in terms of “don’t know” (item non-response). In the case of questions on the field of education, more “don’t know” answers were recorded in the previously implemented CATI of the regular LFS data collection of the 4th wave than in the WEB pilot. 25.7% in case of proxy and 36.1% in case of self-reporting in the CATI 4th wave. The interviewers interference could play crucial part in that particular case.

VI. Conclusions and discussion

28. Future challenges and conclusions of the pilot web data collection. WEB data collection is feasible for the survey, but we propose some adjustments for the data collection and some further improvements of the questionnaire before implementing it in the regular collection of LFS data.

29. Plans for the implementation of the WEB in the regular production of the survey

- The data for LFS in Slovenia in the future will be collected in the mixed mode design. One of the challenges is to integrate WEB data collection in the regular production.
- The current goal is to implement it step by step with conservative approach. Firstly, some smaller and on average younger households will be selected for the WEB mode (e.g. max 2 household members where selected person is between 25 and 39 years old). We will start with the implementation of WEB in the last, 4th wave, since some loyalty to the survey response have already been established from the respondents who have already responded in the previous three waves.
- We still have to analyse the data collected with the pilot in more detail to find out what is the quality of so-called open questions (where respondent have to report their education, occupation and economic activity). In general, compared to CAPI and CATI mode, descriptions given directly by the respondents are of higher quality, accuracy and contain less typos. However, compared to both modes that rely on the interviewer’s interference, WEB turned out to produce more “don’t knows” in some cases, especially with proxy respondents. Nevertheless, collected answers seem to contain enough detail to code them with required precision.

30. Main findings:

- We have not yet integrated the coding list into the questionnaire. We would like to test as a separate action whether implementing complex coding list in the WEB questionnaire is feasible or not. Therefore, we applied for the EU funding of the action to find appropriate way of implementation and to test its feasibility within the Slovenian LFS WEB questionnaire.
- High unit non-response rate. We have achieved around 50% of the responses from the 5th wave. That means that these are our »loyal« respondents and results obtained from the pilot regarding the response rate do not reflect the result that we would typically obtain in the second wave of the data collection “in the field”.
- We experienced very high »Don't know« proportion for some questions. There are several questions where “Don’t know” answers exceeds 5%; both for proxy and self-

- reports. Significant number of “Don’t know” answers were recorded among the open questions, where text box was presented to be filled in with the text or number. Such questions were for examples those on the number of hours usually or actually worked, extra hours worked or number of days of absence from work, as well as questions on the field of education. E.g. 13% of self-reported and 30% of proxy answers to the question on number of hours usually worked at the main job (compared to 3% and 10% in previous CATI mode).
- d. Dependent interviewing worked well, but some questions are too detailed to be answered by proxy respondent. For this reason, it would be desirable that each person in the household would get his or her own questionnaire by e-mail. This scenario would be feasible if we would get e-mail addresses from the Central Population register. The results of the test suggest that we got only 5% of the e-mail addresses.
31. What were the biggest problems and what have we learned from the pilot
 - a. Quite some respondents commented that the questionnaire is too long and that the questions seemed too much alike and are not adjusted to the status of the respondents. For example, for students or even self-employed some questions should be asked differently.
 - b. It is very complex and long questionnaire, where many adjustments were made. Questions and answers were designed in a way that the question is formed different if the person is answering for herself/himself or other household members. This was to reflect the Slovenian grammar.
 - c. Implementation of the coding list for some questions still has to be done and tested.
 32. LFS WEB pilot showed us that respondents are very keen of the WEB surveys. The majority would like to participate in the WEB mode also in the future. This information let us know that at least part of the society is prepared and ready for such WEB questionnaires. Those who responded the pilot survey preferred WEB to CAPI mode.
 33. WEB mode required a totally different approach as with other modes.
 - a. That is why, firstly, we must further investigate, who is the appropriate population for such mode. For this, we need to know who our respondents are and who would rather answer the questionnaire on the WEB. This way WEB mode could improve the overall response rate.
 - b. Secondly, we need to further test the quality of the collected data and find out if WEB mode is suitable for household survey or is it better to adjust it to person survey (some comments obtained from the ‘field’ suggested the latter).
 - c. Thirdly, we must further improve the questionnaire for WEB mode, so the questionnaire will be (i) adjusted to the status of the respondents and (ii) will contain the coding list.
 - d. Furthermore, consideration should be given to how to implement the WEB mode in the regular LFS survey. Which wave(s) to use and how to time it into the data collection that is already very short and does not allow much of extra time to be allotted for the WEB mode. In addition, the new data collection (WEB mode) also requires a new interviewer and data management.

References

Using paradata to manage and monitor multi-mode surveys. Gina Cheung, University of Michigan / CBS - Statistics Netherlands, Cheng Zhou, University of Michigan Marcel van der Steen - CBS - Statistics Netherlands