



Leibniz-Institut
für Sozialwissenschaften

GESIS Panel Online Paradata Documentation

*Kai Weyandt, Bella Struminskaya , Ines Schaurer
Last updated 23 June 2014 (Version 1.0)*

Online Paradata in the GESIS Panel

In the online surveys of the GESIS Panel, paradata are collected on a page-by-page basis. Paradata are the data generated automatically during the response process (Couper, 1998; Kreuter, 2013) and provided unknowingly by respondents (Kaczmirek, 2009). The goal of the collection of online paradata is to study the online response process.

In the GESIS Panel, the paradata are collected by two means: first, via the UCSP (Universal Client Side Paradata) script that was developed by Lars Kaczmirek and second, via the built-in functionality of the online surveying software (Relative Time Stamps, or RTS). The UCSP script allows precise time measurements and collects more information than the built-in software functionality. Both the UCSP and the RTS are described further in detail.

The Universal Client Side Paradata Script

The UCSP script is integrated into the survey software of the GESIS Panel (QuestBack EFS Panel) and is activated on every page of the questionnaire. The content generated by the script (i.e., the information about the answering process) is saved as a string variable. With every page visit, a new string sequence is generated and written at the beginning of the existing string. The information about the first page visit is therefore always at the end of the string. Depending on the number of page visits, the resulting strings can vary in length. For the analysis of paradata, it is necessary to split the string variable into separate variables. The UCSP script is based on JavaScript and demands the activated JavaScript on the side of the respondent. If the respondent has deactivated the JavaScript, no paradata are collected. The UCSP script has several versions. Currently, version 6 is implemented in the GESIS Panel.

Structure and Content of UCSP strings

Overall, 15 variables per questionnaire page are collected. The variable *visits* contains the number of page views (page loads). The variable is filled with information when the "back"-button is used. The variable *dac* contains the information about the activation of the probe „Do-Answer-Check" (i.e., the note that one or several answers on a particular questionnaire page have not been provided by the respondent). The activation of the DAC is not counted as a new page visit. However, currently DACs are not implemented in the GESIS Panel.

The variable *time2firstclick* is the number of milliseconds from page load to the first click on the page. That time reflects the processes of orientation on the page, reading time of the question and time for generating the answers. The variable *time2secondlastclick* measures milliseconds from page load to the second-to-last click. Ideally, the second-to-last click is the last answer, because the last click is the click on the "submit"-button. Thus, this time corresponds to the answering time. The variable *time2lastclick* provides the milliseconds from the page load to the last click. The last click can only be the "submit"-button, therefore, this variable provides the viewing time of the page.

The variable *clickcount* contains the number of clicks the respondents has made on a questionnaire page. The variable *dblclickcount* contains the number of double clicks. The definition of the double click is given by the operating system which the respondent uses. The variables *winx* and *winy* provide the information about the window size or the space which is used for the online survey. *Winx* is the window width measured in pixels, *winy* is the window height in pixels. The variables *scrollxprev* und *scrollyprev* provide the maximum space for scrolling (width and height) for the survey window measured in pixels. Ideally, scrolling width and scrolling height would equal zero.

The variable *msfirstkey* is the time stamp containing the time from page load to the first key stroke. The related variable *mslastkey* measures time from page load to the last key stroke. The variable *cntblur* provides the number of blur-events defined as leaving the survey window. The variable *timeblur* provides the time of the blur-events in milliseconds. The variable *prefill0* contains information about the previous page visits. If the variable does not equal zero, it contains the 15 variables explained above for each page visit.

Below two examples of the structure of the UCSP script and the overview of the variables of the UCSP string (Table 1) are given. The prefix „#6#“ designates the current version of the paradata script. It is added to the string at every page visit and therefore precedes every string sequence. The variables of the last page visit are divided by semicolons. The variables of all other page visits are divided by a slash symbol („/“). The page visits are divided by the version identifier as described above.

Schematic structure of the paradata string with one page visit:

```
#6#"visits";"dac";"time2firstclick";"time2secondlastclick";"time2lastclick";"clickcount";"dblclickcount";"winx";"winy";"scrollxprev";"scrollyprev";"msfirstkey";"mslastkey";"cntblur";"timeblur";"prefill0";
```

An example of the paradata string with two page visits and position numbers:

```
2nd Visit          1st Visit
#6#1;2;3;4;5;6;7;8;9;10;11;12;13;14 #6#1/2/3/4/5/6/7/8/9/10/11/12/13/14/15/16
```

Table 1: Overview of the variables

Position	Label	Description
1	<i>visits</i>	Number of page visits. The first visit has a value of "0", is set to a higher value when the "back"-button is used.
2	<i>dac</i>	One-time activation of the probe (Do-Answer-Check, short: DAC) on the survey page. Currently, probes are not implemented in the GESIS Panel.
3	<i>time2firstclick</i>	Time from page load to the first click in milliseconds (orientation on the page, reading time and answering time).
4	<i>time2secondlastclick</i>	Time from page load to the second-to-last click in milliseconds (time to generate the answers)
5	<i>time2lastclick</i>	Time from page load to the last click, that is the submit button, in milliseconds (page viewing time)
6	<i>clickcount</i>	Number of clicks
7	<i>dblclickcount</i>	Number of double clicks (depends on the system)
8	<i>winx</i>	Window width in pixels
9	<i>winy</i>	Window height in pixels
10	<i>scrollxprev</i>	Maximum width of scrolling in pixels

Position	Label	Description
11	<i>scrollyprev</i>	Maximum height of scrolling in pixels
12	<i>msfirstkey</i>	Time from page load to the first click in milliseconds
13	<i>mplastkey</i>	Time from page load to the last click in milliseconds
14	<i>cntblur</i>	Number of blur-events (left the survey window)
15	<i>timeblur</i>	Duration of the blur-events in milliseconds
16	<i>prefill0</i>	0 if the page has been visited only once; if the page has been visited more than once, the paradata information described above (1-15) is added to the string at the position 16.

Relative Time Stamps collected by the online survey software

The relative time stamps (RTS) are also collected on a page-by-page basis. In contrast to the client-side paradata described in the previous paragraphs, relative time stamps are server-based. That means these time measurements also contain the time it takes for the questionnaire page to load. This time varies depending, for example, on the speed of the Internet connection. Thus, server-side time stamps do not allow the level of precision which can be achieved by the UCSP script.

The relative time stamp for each questionnaire page reports the time difference in seconds between pressing the "submit"-button of the respective page and the absolute time stamp. The absolute time stamp signifies the difference in seconds between 1.1.1970 and the moment of loading the first page of the survey. For the technical details about the logic of the provided timestamp variables please refer to the EFS Survey manual (version 10.1/1.1).

Additional paradata collected by the online survey software

Additionally, the online survey software collects information on the date and time when the respondent first and last accessed the survey, duration of the survey (the difference between those times), the browser used by the respondent, page history (the movements of the respondent throughout questionnaire pages via page IDs), and whether the JavaScript and flash are enabled. For the detailed information on those variables, please refer to the EFS Survey manual (version 10.1/1.1).

Documentation

The paradata variables generated by the UCSP script have the notation of "zs" in the variable names. Relative time stamp variables have the notation of "zr". The paradata variables outlined in the section "additional paradata collected by the online software" are marked by the notation "zp". The paradata variables correspond to the variables containing substantial answers via variable labels. Please refer to the wave reports and the codebooks of the specific waves for specific information on the variable names in each wave.

References

- Couper, M. P. (1998). Measuring survey quality in CASIC environment. Paper presented at the Joint Statistical Meetings of the American Statistical Association, Dallas, TX.
- Kaczmirek, L. (2009). Human-survey interaction: usability and nonresponse in online surveys. Cologne: Herbert von Halem Verlag.
- Kreuter, F. (2013). Improving surveys with paradata. In F. Kreuter (Ed.), *Improving surveys with paradata: Analytic uses of process information*. Hoboken, New Jersey: Wiley.