



WageIndicator

in Argentina, Belgium, Brazil, Denmark, Finland, Germany, Hungary, Italy, India, Mexico, Netherlands, Poland, Spain, South Africa, South Korea, United Kingdom, United States

www.wageindicator.org

THE *WAGEINDICATOR* WEB-SURVEY: EXPLAINING THE DATASET

Kea Tijdens, AIAS/University of Amsterdam, Netherlands
15/01/2007

For downloadable documents related to this text, please visit:

www.wageindicator.org > Research lab > [Downloads](#)

Countries, locales and originator sites

The server assigns the variable **locale**. It indicates the QMS language version of the questionnaire. A country can have one or more locales.

The variable **COUNTRY** refers to the country of the respondent. The data-manager computes if from the locale. The variable label in **COUNTRY** holds both the ISO country code and the name of the country.

The locales, countries and ISO country codes in the WageIndicator survey (end 2006), covering 17 countries

locale	country	ISO code	locale	country	ISO code
es_AR	Argentina	32	it_IT	Italy	380
nl_BE	Belgium Dutch spoken	56	es_MX	Mexico	484
fr_BE	Belgium French spoken	56	nl_NL	Netherlands	528
pt_BR	Brazil	76	pl_PL	Poland	616
dk_DK	Denmark	208	en_ZA	South-Africa	710
fi_FI	Finland	246	kr_KR	South-Korea	410
de_DE	Germany	276	es_ES	Spain	724
hu_HU	Hungary	348	en_GB	United Kingdom	826
en_IN	India	356	en_US	United States of America	840

Two variables indicate the site from where the questionnaire was completed. The first one is **orgsit2**, indicating whether the questionnaire comes from a short, a long or a category questionnaire in the country of survey. The second variable is **orgsit3**, indicating the url of the website where the respondent has completed the questionnaire.

Date and time meta variables

A number of data and time variables are assigned during survey completion, as shown in the Table. The servers register the local server time and date for every respondent. Based on **COUNTRY** and **REGIHOME** data, the data-manager runs a program to compute the respondent's local time and date, when the completion of the questionnaire was started and ended. **survetb** **survete** **survedat** On these three variables all other variables presented in the Table are based.

Metadata variable names and labels

variable name	format	variable label
survetb	hh:mm	Time of survey (begin) server time USA GMT-6, NL GMT+01, IND GMT+5.5
survete	hh:mm	Time of survey (end) server time USA GMT-6, NL GMT+01, IND GMT+5.5
survedat	dd.mm.yyyy	Date of survey, based on survetb
SURVEYY	Numeric	Year of survey
SURVEMM	Numeric	Month of survey
SURVEWW	Numeric	Week of survey
SURVEDD	Numeric	Day of survey (local GMT Standard Time)
SURVEHH	Numeric	Hour of survey (local GMT Standard Time)
SURVTIME	Numeric	Minutes needed to complete survey
SURVEDW	Numeric	Day of the week
SURVEDM	Numeric	Day of the month
SURVEDY	Numeric	Day of the year

Number of survey version and release

SURVENR is the variable indicating the version number of the questionnaire, whereby the 5 indicates the year 2005 of the version and the 2 indicates the version in this year. The questionnaire in the Netherlands started with survey number 1 to 41. For the international data, version 42 is the first.

The data of the continuous web survey is released on a quarterly basis. The release is numbered from 1 onwards. It is strongly advised that any analyses should be preceded by a cross table of **RELEASE** and the variables under study. Any changes in the questionnaire, either due to changes in the QMS or due to technical failures, can be checked this way.

The variable **storage** indicates whether the questionnaire is fully submitted or not completed due to time-out. This variable is available since **RELEASE**=5. The dataset does NOT include cases with time-out. The dataset with incomplete questionnaires is available on request.

ID number

Each observation in the dataset has two identification numbers. The variable **idkey** is a unique number, automatically assigned once a respondent has started to complete the questionnaire. It has 20 characters, both numerical and alphanumerical. It is among others is used to identify the panel members. The variable should NOT be used for sorting.

The data-manager assigns **IDNR**. When sorting the dataset, **IDNR** should be used. The Table shows the first and last **IDNR** per release.

*The first and last **IDNR** per release.*

RELEASE	Release number	IDNR Minimum	IDNR Maximum
1	2004/4	4208710	4246491
2	2005/1	4210266	4275372
3	2005/2	4275374	4319648
4	2005/3	4319649	4340443
5	2005/4	500000003	500054300
6	2006/1	510600001	510641913
7	2006/2	510700001	510735015

Variable names in capital

In the dataset, new variables are computed. For identifying these variables, capital letters are used. For example **AGE** is a variable derived from **SURVEYY – yybirth**. Here, **AGE** and **SURVEYY** are created after data-intake and **yybirth** is direct data-intake, notably in the question 'In what year are you born'. Parallel questions aiming at various target groups are computed into one variable, in order to minimize the number of variables. Similarly, the original data from a search tree, whereby each click is stored, is computed into one variable.

Identifying variable names

The variables in the dataset all have variable names and labels, and so have the values. The first letters of the variable names refer to groups of questions.

Variable names and labels

first letters in var name	variable names refer to	first letters in var name	variable names refer to
break	spell out of the labour force	lcc	child care arrangements
cao	collective bargaining agreement	mem	member of
chld	child	occ	occupation
cob	country of birth	pa	participation in firm's schemes
comm	commuting	pref	preferences
comp	computer	regi	region
con	working conditions (health and safety)	sat	satisfaction
cont	contract	sea	search behavior when looking for jobs
dep	department	seek	items important when looking for jobs
ddwor	last work day, last work week	surv	survey
edu	education	supv	supervising/supervisor
firm	firm	ti	timing of work
fr	fringe benefits	vol	voluntary work
hh	household characteristics	wa	amount in case of fringe benefit
hrs	hours of work	wage	wage
job	job	yy	calendar year

Variable formats

The dataset employs three variable formats, i.e. numeric, string and date/time. By far the most variables have the numeric format with widths ranging from 2 to 12 numbers and from 0 to 2 decimals.

The string format is used for the open-ended questions in the questionnaire, notably for the web-visitors' comments on their industry, their occupation, on the payment period of their wage, and at the end their comments on the questionnaire. Apart from that, a few metadata-variables are in the string format, such as **locale**, **orgsit2** and **orgsit3**.

The date/time formats are used in the form of DD:MM:YYYY for days or HH:MM for hours:minutes. Variables with a HH:MM format are recoded into numerical values (see Table).

Variables in TIME and NUMERICAL format

time	39:00	39:15	39:30	39:45	40:00
numerical	39.00	39.25	39.50	39.75	40.00

Calendar years

Questions relating to the dates of previous events are asked in calendar years using a drop-down menu. In some questions, a drop-down menu allows for a choice of the 12 months in addition to the calendar years. The recent year in the drop down is updated every year. Thus, in 2007, the variable **yybirth** (year of birth) runs from 1997 (=survey year – 10) to 1930, and **yycuempl** (year joining current employer) runs from survey year to 1950. The starting years 1930 and 1950 do not change over time.

Note that the calendar years are updated each year. So, don't expect the value 2005 for the variable **yycuempl** when looking at the data of the 2002 survey!

Text boxes

The web survey has four open-ended questions, as can be seen in the Table. Except for the variable **survetxt**, the text boxes are available in the dataset.

Four open-ended questions

var name	question
sectext	If you want to specify the main business activity of the organization where you work more accurately, please do so here
occtext	If you want to specify your occupation more accurately, please do so here
wagetext	If you want to specify your payment period, please do so here
survetxt	If you have any comments on the questionnaire, please do so here