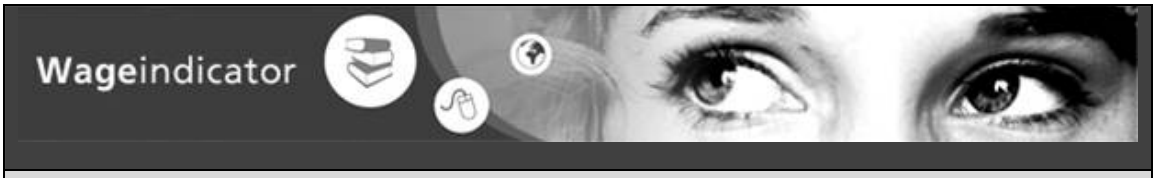

in Belgium, Brazil, Denmark, Finland, Germany, Hungary, Italy, India, Netherlands, Poland, Spain, South-Africa, South-Korea, United Kingdom, United States

Introduction to the *WageIndicator* dataset

Report Version: Final

Date: 09.12.2005

Author: Kea Tijdens, AIAS, University of Amsterdam, NL

NL - www.loonwijzer.nl/
DE - www.lohnspiegel.de/
FI - www.palkkalaskuri.fi/
ES - www.tusalario.es/
PL - www.twojezarobki.pl/
IT - www.iltuosalario.it/
BE - www.loonwijzer.be/
BE - www.votresalaire.be/
UK - www.paywizard.co.uk/
DK - www.lontjek.dk/
DK - www.lontjek.com/
WW-www.wageindicator.org/

TABLE OF CONTENTS

COOPERATING PARTNERS IN *WAGEINDICATOR* ARE:

BE	HIVA; ACV-CSC;	Leuven; Brussels
BR	DIEESE ; all trade unions in Brazil;	Sao Paulo
DE	WSI; Verdi, DGB and other trade unions in Germany;	Dusseldorf; Berlin
DK	NewInsight; SID;	Copenhagen
ES	University of Salamanca; CC.OO; UGT; Comfia; Infojobs.net;	Madrid; Salamanca
FI	Ttopalvelu Käyttötieto Oy; SAK, Akava, STTK;	Helsinki
HU	MSZOSZ; SZGTI; ESZT; Nők a Holnapért Alapítvány; V2 Excelsior Bt;	Budapest
IN	ITPF India, IIM Ahmedabad; NIIT; Naukri; Institute of Public Enterprise Hyderabad; ICFAI Channai; Indian Institute of Science Bangalore;	Ahmedabad; Bangalore; Chennai; Hyderabad
IT	CHILD/ U Torino; Cgil;	Rome; Torino
KR	KLSI; all trade unions in South Korea,	Seoul
NL	Stichting Loonwijzer/WageIndicator Foundation; University of Amsterdam AIAS; FNV; Monsterbaord.nl;	Amsterdam
PL	Universty of Economics Poznan; Solidarnosc; Gazeta.pl/Gazeta Wyborczej;	Poznan; Warsaw
UK	IDS; TUC;	London
US	Worklife Program of Harvard Law School;	Boston
ZA	Naledi; Cosatu and all trade unions in South Africa;	Johannesburg; Capetown

FUNDS FOR *WAGEINDICATOR* COME FROM:

Netherlands	Ministerie van Sociale Zaken; FNV Mondiaal; NWO (Data-Archive project 470-00-400).; Task Force ouderen;
European Union	FP6-2004-506590; EQUAL; ESF-3

SME'S WORKING FOR THE WAGEINDICATOR PROJECT:

Webword	Project & Web management, www.webword.nl
Wyldebeast & Wunderliebe	Web programming, www.wyldebeast-wunderliebe.com
IfThen	Web design, www.ifthen.nl
Datamatch	Datamanagement, www.datamatch.org
Zodra	Web journalism, www.zodra.nl

1. INTRODUCTION

In 2000, the *WageIndicator* project started developing a plan for a paper-and-pencil survey, aiming to use the survey data for establishing a website with salary information for women's occupations. The survey was published in the three largest women's magazines in the Netherlands. By the end of 2000, the survey was put online on a website associated to one of the women's magazines. This turned out to be a great success, because approximately half of the almost 15,000 questionnaires were completed online. In 2001, the first *WageIndicator* website went online with the *Salary Check* and the web-survey, that from then on also aimed at the male work force. In the years to follow, dedicated websites for youth workers, for workers over 40 and for the self-employed were initiated. Several features were added to the website to increase its attractiveness, such as a Minimum Wage checker and a time-budget profile for age-gender groups.

Initially, the *WageIndicator* website was funded from small funds, a prize, help of large media, and volunteer work. In 2004, the survey could expand to Belgium, Denmark, Germany, Finland, Italy, Poland, Spain, and United Kingdom, thanks to the 3-yr EU-FP6 grant for the WOLIWEB project (FP6-2004-506590). In 2005, websites could start in Brazil, India, South Africa, and South Korea, thanks to a 3-yr grant of the Dutch fund for Development Aid for the GLOBAL project. By the end of 2005, websites could start in the USA with own funds and in Hungary with funds from a EU-EQUAL project. For 2006, expansion to Argentina and Mexico is funded from Development Aid and expansion to Turkey with funds from the Dutch Ministry of Labour. In 2005, the Dutch scientific council NWO funded the merger of the *WageIndicator* datasets 2000-2003 and 2003-2004.

Each country has its own team of researchers and web-managers and its own *WageIndicator* website, consisting of job related content, an anonymous questionnaire with a prize incentive, and a free and crowd-pulling *Salary Check* presenting average wages for occupations, based on data from the questionnaire. Additionally, the project includes web-marketing, publicity, and answering visitors' email. Coalitions with media groups and publishing houses with a strong Internet presence exist.

Monthly, the *WageIndicator* website in the Netherlands receives over 300,000 unique visitors, and is among the highest ranking dedicated websites in the country. Approx. 1.5% of the visitors completes the questionnaire. As for the 7 countries, more than 92,000 individuals from completed the questionnaire from January to August 2005.

This paper introduces the *WageIndicator* dataset. More information can be found at www.wageindicator.org/, section Research Lab, where downloadable research papers, codebooks, classification data, and other information related to the survey are posted. The Lab has also a syntax shop, where SPSS programs can be downloaded to add value labels or compute new variables. Access to the data is documented in the *WageIndicator Foundation*'s Data policy paper, posted at the Research Lab.

2. THE TOPICS COVERED IN THE QUESTIONNAIRE

The *WageIndicator* questionnaire aims to collect information on wages and working conditions in the labour force. The questionnaire is organised into six clusters, and each cluster ends with a set of attitude items (see Table). The reasons for including these clusters of questions are explained elsewhere (see project website).

Table. Six clusters in the Wage Indicator questionnaire

cluster	Topics
A	YOUR OCCUPATION
	education, occupation, industry, training
B	YOUR PLACE OF WORK
	firm characteristics, branch, firm size, percentage female, region, workplace characteristics – staffing levels, cooperation, division of labour IT-use at the workplace, and attitudes towards IT-adaptation
C	YOUR EMPLOYMENT HISTORY
	employment record – years of experience, career break, job search
D	YOUR WORKING HOURS
	working hours, overtime, timing of work, and working time preferences
E	YOUR EMPLOYMENT CONTRACT AND SALARY
	employment contract, wages, fringe benefits, bonuses, and wage perceptions
F	PERSONAL QUESTIONS
	individual characteristics – age, gender, ethnic background, region, household composition – marital status, children's age, division of household chores, job and life satisfaction

The *WageIndicator* questionnaire uses international renowned classifications for measuring industry, occupation, country and region of residence, trade union, and collective agreement. Survey questions about these variables are most commonly asked either as an open text field or from a short aggregated list. For web-surveys with large amounts of competed questionnaires, we have developed the so-called choosers, as these are the only workable solution for gathering detailed information without being restricted to either a huge recoding effort or a limited list of items. A chooser is a two or three-tier ticking list, detailing an aggregated level in the first tier to a narrow level in the second or third tier. The chooser technique in the *WageIndicator* questionnaire allows the web-visitor to go easily back-and-forth. The spreadsheets with the choosers can be downloaded.

3. TARGET POPULATION AND ROUTING

The target population of the *WageIndicator* questionnaire consists of individuals in paid employment and job seekers, i.e. the labour force. It excludes individuals not attached to the labour market, such as housewives, pensioners or persons performing voluntary or unpaid work. The questionnaire aims to include all forms of waged employment. Apart from workers in dependent employment it aims to include apprentices, employers, own-account workers, freelancers, workers in family businesses, workers in the informal sector, partly unemployed, disabled or retired workers, and school pupils and students with a job on the side. If the questions in the survey only address the main group in the labour force, the atypical groups are more likely to drop out during questionnaire completion. Therefore, all questions have been reviewed as to the extent that they address all groups. Where necessary, new questions have been added, addressing the atypical groups.

Compared to paper-based surveys, web-surveys are advantageous as they allow for a sophisticated routing through he questionnaire. In the *WageIndicator* survey, the routing technique is used to address the atypical groups as much as possible. For five groups, the questionnaire has a unique routing through the questionnaire (see Table). Selection for the routing is based on the first question: 'Which description matches best your current employment activity?'. The variable is called CONTST.

Table. Routing based on current employment activity (CONTST).

value	CONTST labels	routing	explanation
40	Employee	A	Own routing
30	In job creation scheme	A	Similar routing as 1 Employee
2	Self-employed, own-account worker	B	Own routing, particularly for the questions about wages
3	Family worker	B	Similar routing as 3 Self-employed
25	Apprentice/ trainee	C	Own routing, particularly for the questions about education
20	School pupil, student in full-time education, even if on vacation	D	Follow-up question: Do you have a paid job? If no: ALERT THIS QUESTIONNAIRE DOES NOT AIM AT YOUR GROUP. If yes: Special routing, particularly for the questions about education
10	Unemployed / looking for a job	A	Follow-up question: Did you ever have a paid job? If yes: ALERT COMPLETE QUESTIONNAIRE FOR YOUR LAST JOB. Similar routing as 1
		E	If no: Special routing for individuals who never have had a job
15	Sickness benefit / incapacity for work	A	Did you ever have a paid job? If yes: ALERT COMPLETE QUESTIONNAIRE FOR YOUR LAST JOB Similar routing as 1 Employee
		E	If no: Similar routing as 10 Unemployed
5	Free for task	B	Own routing, particularly for the questions about wages
4	Contract by results	B	Own routing, particularly for the questions about wages
50	Other	A	Similar routing as 1, Employee

4. THE QUESTIONNAIRE MANAGEMENT DATABASE (QMS)

The worldwide *WageIndicator* web-survey is managed in the Netherlands. The websites are hosted on three servers in the Netherlands, the USA and India. The survey has a sound, multilingual Questionnaire Management System that was totally renewed in the early half of 2005 to facilitate the worldwide use of the questionnaire.

The QMS is implemented in a Plone environment, using a Zope/Python based CMS. The QMS consists of a maintenance module for the datasets, a maintenance module for the presentation layer and a module for the selection process (the so-called chooser). The implementation was based in an Eclipse environment and is based on Java, Struts, JSP and Maven. For the chooser application a management web application was build using Struts and Hibernate. The application uses tomcat with a MySQL database. The Socrates questionnaire engine is an Open Source project. The engine is extensively tested, both by the current research team as by the public at large, visiting the website. The QMS has a codebook mode, presenting the content of the presentation layer, i.e. the questionnaire, except for the content of the choosers.

The engine facilitates routing by means of relevance rules. It has obligatory questions, has an option to switch questions and answer categories on/off per country and an option to move questions through the questionnaire. Finally, in case of matrix questions the engine allows to draw random items from an item pool. The QMS assigns system missing values to all variables derived from questions, answer categories and items that are not shown to the respondent.

The data is stored on a special, secure website, accessible to the data-manager only. The data is stored as delimited text. The data-manager taps the data usually once a week, and converts the delimited text into a SPSS statistical file. The data-manager assigns user missing values to missing values that are not system missing values. The data-manager assigns user missing or zero values to checkboxes.

Even multi-country surveys require country specificity of the questions. For example, the question about commuting asking for 'Distance in kilometers (one way)' is switched off for the UK and the USA, whereas the question 'Distance in miles (one way)' is switched on for the two countries. The question 'Do you have a mini-job?' is switched on only for Germany, as this phenomenon does not exist in other countries. In the second week of January 2005, Germany, UK and to a minor extent Spain and Poland expressed their desire to reduce the number of questions,

because visitors had complained that the questionnaire was too long. Therefore, some questions were switched off for some countries. Although the codebook shows per release which questions are switched off for which countries, researchers analysing the dataset, are advised to make a cross tabulation of a target variable with country and survey week to check the data. By the end of 2005, for special occasions, i.e. links in websites of large media partners, a short questionnaire was needed, addressing workers in waged employment only and asking the obligatory questions only. A variable will indicate whether the short or long questionnaire was taken.

5. DATA INTAKE AND DATA QUALITY

In August 2005, data-intake from seven countries varied from 100 to 500 completed questionnaires per day. In a year from now, the data-intake is expected to double. Since 31.09.2004, the dataset is released on a quarterly basis. The data-manager taps the data usually once a week and tests the data with each tap as for completeness. This is sufficient when an observation has values for six critical variables, notably education, occupation, industry, wage, sex, and year of birth. In addition, this test aims to detect any technical failures.

As for uniqueness, the data manager checks the data per release. This test identifies respondents that have completed the questionnaire multiple times. This is the case for observations with the same value on the sum of the values for region of work, industry, occupation, year of birth, gender, education, presence of children and gross wage. Double cases are excluded.

Out of range values seem to be impossible in a web-survey. Nevertheless, they occur. A major reason is found in the data from textboxes. Particularly when visitors have typed a semicolon in the textboxes, when converted from the txt-file to the SPSS-file the semicolon is interpreted as a separator. This causes cell-overflow, i.e. the data is not stored in the proper cell, but in the next cell. The *WageIndicator* programmers have by now solved the problem to a large extent. The dataset is cleaned for out of range values.

Once the *WageIndicator*survey starts, a country mostly uses a couple of weeks to test the web-mode of the questionnaire. Though this data is stored, the data-manager deletes this data.

Table. Week of first data-intake after the test period

Country code	Country	Week 2004	Day 2004	Week 2005	Day 2005
DE	Germany	44			
ES	Spain	45	1		
UK	United Kingdom	48			
BE	Belgium	49			
PL	Poland	50			
DK	Denmark			13	5
FI	Finland			16	6
IT	Italy			not known yet	-

In summary, the tests for data quality include missing values on six main variables, out-of-range values, observations in the test period, and duplicate cases. For the data from 31.09.04 to 14.09.05, in total 2.7% of the cases were invalid and

therefore deleted from the dataset. Taking into account the improvements of the QMS, lower percentages are expected for the near future.

Until November 2005, only fully completed questionnaires were registered, i.e. from the respondents that had pressed the button SEND at the end of the questionnaire. From November onwards, incomplete questionnaires are also registered as long as visitors have ticked at least the first question. A variable will indicate whether the data comes from completed or from incomplete questionnaires.

6. METADATA

The WageIndicator data include several metadata variables (see Table).

Table. Variable names and labels of the metadata

var name	var label
IDNR	ID-number (1000000 * SURVENR) + ID
id	ID-number per version
SURVEDAT	Date of survey (Date format)
SURVEDD	Day of survey (1=Sun, .. , 7=Sat)
SURVEHH	Hour of survey (1 – 24)
SURVEMM	Month of survey (1 – 12)
SURVEWW	Week of survey (1 – 53)
SURVEYY	Year of survey
survetb	Time of survey (begin) (Time format)
survete	Time of survey (end) (Time format)
SURVTIME	Minutes needed to complete survey
orgsite	Originator-site
locale	Locale
SURVENR	Version of survey
RELEASE	Release number
COUNTRY	Country of survey (ISO)

As for the variables id and IDNR, each observation in the dataset has two identification numbers, whereby 'id' is automatically assigned once the questionnaire is stored in the QMS data-interface. The data-manager assigns 'IDNR', which is (1000000 * SURVENR) + ID. In case of sorting the data, please use always IDNR.

As for the variables SURVENR and RELEASE, the former is a variable indicating the version number of the questionnaire, whereby the 4 indicates the starting year 2004 of the version and the 2 indicates the version in the starting year. For the international data, version 42 is the first (see Table). For the Dutch Loonwijzer dataset version numbers run from 01 to 41. This paper is written after the fourth quarterly release of version 42, covering the period until September 9, 2005.

Table. Versions of the Wage Indicator Questionnaire

version	dates	name	mode	remarks
01	000-10 / 2000-11	Women's Wage Indicator	paper	women only
02	2000-11 / 2000-12	Women's Wage Indicator	web	women only
11	2000-12 / 2001-04	Women's Wage Indicator	web	women only
12	2001-05 / 2002-01-22	Wage Indicator	web	women and men
21	2002-01-23 / 2002-10-25	Wage Indicator	web	adaptation to Euro + few minor changes
22	2002-10-26 / 2003-02	Wage Indicator	web	seven extra questions for employees in health care
31	2003-02 / 2003-12-16	Wage Indicator	web	better questioning of industry, occupation and collective agreement + minor changes
32	2003-12-16 / 2004-04-29	Wage Indicator Time data	web	extra time-budget questions
41	2004-04-25 / 2004-10-16	Wage Indicator Time data	web	extra time-budget questions
42	2004-10-16 / current	Wage Indicator	web	improved questioning of industry, occupation and collective agreement

The dataset includes a variable COUNTRY, based on the ISO country classification. The variable locale specifies the language of the questionnaire in connection to the country.

Table. The locales, countries and ISO country codes in the WageIndicator survey

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

Since May 2001, the month of survey is included in the dataset, with the exception of the months 3-11 in the year 2003. Since early 2002, the day of survey is recorded, with the exception of the months 1-11 in the year 2003.

Since 29.12.2004, the *WageIndicator* data includes both the starting and the ending time of the completion. The variable SURVTIME indicates the time needed to complete survey, i.e. survete – survetb.

7. VARIABLE NAMES AND FORMATS

The first letters of the variable names refer to groups of questions.

Table. 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 behaviour 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

7.2 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. Parallel questions aiming at various target groups are computed into one variable (see Table), in order to minimise the number of variables.

Table. Recode of parallel questions aiming at various target groups

[Rule!=NOT FOR School pupil, student in full-time education] AND [Rule!=NOT FOR Never had a job] AND [Rule!=NOT FOR Apprentice, trainee]			
A32	edujobtr	How much time is needed to become fully effective in your job for someone with your qualifications?	-> EDUJOBTR
		☐ 0 No training period required	
		☐ 1 1 - 2 days, etc	
* [Rule=FOR School pupil, student in full-time education]			
A33	edujobt 1	How much time is needed to become fully effective in your job?	-> EDUJOBTR
		☐ 0 No training period required	
		☐ 1 1 - 2 days, etc	

7.3 VARIABLE FORMATS

The dataset has three variable formats, i.e. numeric, string and date. 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 textboxes in the questionnaire, notably for the web-visitors' comments on their occupations and on the payment period of their wages and at the end their comments on the questionnaire. These three variables are called occtext, wagetext, and survetxt. Apart from that, a few metadata-variables are in the string format, such as locale and orgsite (Originator-site). The date formats are used in the form of HH:MM for hours:minutes and in the form of DD:MM:YYYY for the date of survey for the variables listed in the Table.

Table. Variables in TIME and DATE format

	Variable name and label	Minimum	Maximum
hrscare	Hours a week spend on unpaid care	1:00	80:45
hrscont7	Minimum working hours per week	1:00	80:45
hrscont8	Maximum working hours per week	1:00	80:45
hrscontr	Working hours according to contract per week	1:00	80:45
hrsfirm	Standard fulltime working week	1:00	80:45
hrshelp	Hours of paid domestic help per week	1:00	80:45
hrshkeep	Hours of housekeeping per week	1:00	80:45
hrsreal	Working hours regular per week	1:00	80:45
hrssecon	Working hours in second job per week	1:00	80:45
hrsvolun	Hours of voluntary work per week	1:00	80:45
hrswage	Waged hours per week	1:00	80:45
hhhrspa	Working hours partner per week	1:00	80:45
survetb	Time of survey (begin)	0:00	23:59
survete	Time of survey (end)	0:00	23:59
SURVEDAT	Date of survey	30.09.2004	14.09.2005

The variables with the hours:minutes format are recoded into numerical values, as presented in the Table.

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

7.4 CALENDAR YEARS AND MONTHS

Questions relating to the dates of previous events are asked in calendar years using a drop-down menu (see Table). In some questions, a drop-down menu allows for a choice of the 12 months in addition to the calendar years. In 2005, the variable yybirth (year of birth) runs from 1995 (=survey year – 10)-to-1935, and "year of joining the current employer" runs from survey year – to - 1950. The years 1935

and 1950 do not change over time. The Table shows the list of calendar years present in the dataset.

Table. Calendar years in the dataset

Var name	Var label	Condition
yyarriva	Year arrival	If not born in COUNTRY OF SURVEY
yybirth	Year of birth	All
yybreak	Year stop with salaried employment	If break > 3 months or > 1 yr
yyrenter	Year return to salaried employment	If break > 3 months or > 1 yr
yycuemp1	Year starting own business	If self-employed/family worker
yycuempl	Year with current employer	If employee
yycuposi	Year entering current position	If employee and >=2 positions with current employer
yyeduca	Year finishing full-time education	All, except school pupils and students
yynojob	Year stopped working in last job	If currently unemployed/disabled
yyfstjo1	Year entering apprenticeship	Apprentices
yyfstjob	Year entering salaried employment	All
yyfullti	Year started working fulltime	If changed from part-time to full-time
yypartti	Year started working part-time	If changed from full-time to part-time
yyolchld	Year oldest child born	If children
yyyochld	Year youngest child born	If children

8. CLASSIFICATIONS USED

The *WageIndicator* questionnaire uses international renowned classifications for variables such as education, ethnic background, industry, occupation, country and region of residence. A previous paper details how these classifications are used and how we have developed the so-called 'choosers', facilitating the web-visitors' searching for the applicable, detailed item of the classification. A chooser consists of a two- or three-tier ticking list from an aggregated level in the first tier to a detailed level in the second or third tier. Underlying any chooser, a database links first tier items to second and to third tier items.

A chooser must meet contradictory demands. User-friendliness requires a minimum number of words, and thus a minimum number of items and tiers, as it reduces the likelihood that visitors quit the questionnaire before the end. Yet, visitors must be able to identify their particular enterprise or occupation, and may not be able to identify broader concepts. It should be avoided that visitors must think whether a hairdressing enterprise is part of the retail sector. Therefore, the chooser requires detailed lists of items, which is also preferred for data-analysis. For reasons of user-friendliness each list in the chooser, be it in the first, second or third tier, is automatically sorted by alphabet for each language.

As for the industry code, the variable NACE4NUM specifies he NACE industry classification. This variable includes all countries except the Netherlands (SEC5DIGT) and Germany (SECTOR_D). A syntax file will become available for recoding these industries into the NACE4NUM variable.

As for the occupation code, the variable ISCO9NUM specifies he ISCO occupation classification. This variable includes all countries except the Netherlands. This variable includes all countries except the Netherlands (OCC7DIGT) and Germany (OCCUPA_D). A syntax file will become available for recoding these occupations into the ISCO9NUM variable. Recoding the occupations in the Netherlands to the ISCO9NUM variable may however take some time.

As for currencies, in most countries wages are all paid in the same currency. In some countries, such as Poland, wages may be paid in various currencies. For the latte category of countries, a question asks 'In which currency did you receive your last wage?'. In the dataset, a variable WAGECUR is created, including the currencies of the reported wages and the currencies of all countries where this question is not asked. PPP's are not yet available for the dataset.

Table. WAGECUR Currencies used in the WageIndicator data.

Country code and currency	Country code and currency
1 PL Zlotty	8 ZA rand
2 EURO	9 AR Peso
3 GB Pound	10 HU Forint
4 US Dollar	11 KR Won
5 DK crowns	12 MX Peso
6 BR cruzeiro	13 TR Lira
7 IN rupees	

As for measuring country and region the ISO-country codes and the NUTS-region codes have been used. The latter classification is only available for European countries. For countries outside the European Union, an own classification will be used for region. The following variables include country or region codes. In the Netherlands, the numbers of the postal code are asked. In the dataset, these are recoded into region.

Table. Variables with country and region data.

Var name	Var lable	Var name	Var lable
FIRMOWCO	Country of foreign investment	REGIHOME	Region home address
COBSELF	Country of birth self	REGIWORK	Region workplace
COBMOTHE	Country of birth mother	REGIBIRT	Region of birth
COBFATHE	Country of birth father		

As for education the ISCED classification is used to recode the national education levels according to the question: 'What is the highest level of education you have attained (with certificate)?'. The variable EDUCAT presents the national specific education. It is a variable with the educational categories 1-20 reserved for Netherlands, 21-60 for United Kingdom, 61-80 for Finland, and so on. This variable allows for country specific analysis, using the national education data. Education is measured as the highest level of education, unless the respondent is a school pupil or student in full-time education. In that case the current education is measured. If respondents have in the first question indicated that they are a school pupil or student, the education question is asked differently. The two variables are recoded into one variable EDUCAT.

The dataset has also a variable EDUISCED (ISCED education level) that recodes the national education categories into the worldwide ISCED classification.

*****

9. RELEVANT PUBLICATIONS

Andralojc, Magda & and Paulien Osse, Relation between Completed questionnaires And webvisitors of WageIndicators In Woliwebcountries, Report for web managers meeting June 17–19 2005

Tijdens, Kea & Cecile Wetzels, 2005, Classifications and the chooser-technique used in the *WageIndicator* questionnaire. WOLIWEB paper, downloadable from www.wageindicator.org, section research lab.

Tijdens, Kea, 2004, The dataset, measurement issues and the methodology of the Dutch Wage Indicator Internet Survey. Amsterdam University of Amsterdam AIAS Working Paper WP25, downloadable from www.uva-aias.net/files/aias/wp25.pdf