

Measuring and coding occupations in multi- country surveys

AIAS

Amsterdam Institute for
Advanced labour Studies
University of Amsterdam

*DASISH, Mannheim,
THU 13 December 2012*

*Kea Tijdens
University of Amsterdam, NLD*

- 1. The state-of-the art: measuring and coding occupations**
- 2. The WISCO occupation database for WageIndicator web survey**
- 3. Empirics about occupations**
- 4. The way ahead: an online measuring and coding tool**

- **Projects**

- scientific coordinator WageIndicator web survey (since 2001):
-> volunteer, continuous, multilingual survey on work and wages posted on national WageIndicator websites in 75 countries
- coordinator EurOccupations EU-FP6 2006-2009: building an 8-country occupations database for socio-economic research
- member of the European WEBDATANET (COST Action IS1004 2012-2016) about web-based data collections
- work package coordinator INGRID EU-FP7 2013-2016: tools for measurement of occupations & tools to detect and classify new jobs and new skills



1) The state-of-the art: measuring and coding occupations



The challenge

- **The national stocks of job titles ...**
 - are large: 10,000's – 100,000's of job titles per country
 - are unstructured: 'fuzzy text'
 - are dynamic: many entries and exits over the years
 - but .. comparable across Europe (relative few country specific titles)
 - the challenge: to classify these job titles into the 433 ISCO-08 occupational units and to do that consistently across countries

Hierarchy	#
ISCO-08 1-digit (skill levels)	10
ISCO-08 2-digit	42
ISCO-08 3-digit	131
ISCO-08 4-digit	433
Self-identification database (WISCO)	1,800
Coding indexes	5-50,000's
Job titles	10-100,000's

Answers to the survey question

	Open ended question (OEQ)	Closed question (self-identification)	TSM Text string matching (self-identification)
PAPI	√	√ (max 25)	--
CATI	√	√ (max 5)	--
CAPI	√	√ (max 50 in showcard)	--
CAWI	√	√ (max 2500)	√
coding needed	√	-	-
required	coding index	occupation database + search paths	occupation database
used by	most surveys	employment agencies, job sites, few surveys, WageIndicator web survey	WageIndicator web survey (feb 2013)



What do respondents do?

- **Respondents do know ...**

- their job title from workplace, job evaluation, or job description
- and they are mostly proud and eager to tell
- -> very few 'don't know' and 'don't want to say' answers

- **Respondents do not know ...**

- the titles of ISCO-08 4 digit occupational units
- what kind of answers the survey holder is looking for (some surveys solve this problem with examples, e.g. school teacher)
- -> crude answers -> aggregation heterogeneity (ISCO 1–4 dgt)
- -> ambiguous, irrelevant answers, or abbrev. -> no coding
- in CAWI & CATI interviewers can correct,
- in PAPI and CAWI no correction, of these CAWI performs worst



Can respondents self-identify?

- **Self-identification ...**

- when using highly aggregated lists of occupations
 - > aggregation bias: respondents do not fit their job titles consistently into highly aggregated categories
- when using disaggregated lists of occupations
 - > the smaller the distance to their own job title, the better they are able to classify their job title into an aggregated category

- **... only in web surveys**

- in WageIndicator web survey self identification of job title into a list of 1,800 occupational titles (using the WISCO database)
- disaggregated lists of occupations only possible in web surveys (maybe in CAPI by respondent or by interviewer, no experience)



Coding Open Ended Question

- **Single-country coding**

- requires index, coding into a classification -> costly & time consuming
- survey agencies use (semi)-automatic coding indexes (in various stages of development), using previously coded text strings
- sharing coding indexes across agencies is not common
- recent problem: some agencies use their ISCO-88 indexes with cross-over tables avoiding the investments to build a new index

- **Multi-country surveys -> black box coding**

- multi-country coding indexes do not exist (to my knowledge)
- -> in multi-country surveys the survey holder does not know whether occupations are coded similarly across countries
 - > coding is to some extent a black box
- -> need for multi-country, publicly available coding indexes



2) The WISCO occupation database in the WageIndicator web survey

WageIndicator web survey

- **WageIndicator web portal**

- national websites in 75 countries in 5 continents (end 2012):
 - *Salary Check*: wages earned per occupation using WISCO database
 - information on labour law, minimum wages, vacancies, education, etc
- started 2001 in NLD, since 2004 to EU, since 2006 beyond EU
- millions of visitors: job holders, job seekers, school pupils
- all websites invite visitors to complete a volunteer web survey on work and wages in return to free information provided (lottery prize)

- **WageIndicator web survey**

- multi-country, continuous web survey in national language(s)
- survey question ‘What is your occupation?’ is no OEQ
 - > coding too costly + not-identifiable or crude answers,
- self-identification in 3-step search tree using WISCO database
 - > multi-page filtering: each step on a new page, with back-and-forth

WISCO World Database of Occupations

- **History occupation database**

- initial occupation database coded in classification Stat. Netherlands
- in 2004 NL occupational titles translated into English source labels and assigned ISCO-88 codes
- some new WageIndicator countries translated this source list
- others send their lists of occupations, mostly coded ISCO-88
 - > we could not compare these occupational titles across countries
- in 2006 database had > 10,000 titles, many only for one country

- **WISCO database of occupations**

- EurOccupations project (2007-09):
 - > harmonizing the database for 8 EU countries
 - > adapting to ISCO-08
 - > one search tree for all countries
- WageIndicator : translating the database for other countries



WISCO database and search tree

- **The database**

- approx. 1,800 occupational titles, coded according to ISCO-08 (5 digit)
- all occupational titles have English source labels
- translated and checked for 75 countries (23 in EU)
- a 5-digit ISCO-08 code refers to the same occupation in all countries
- national specific occupational titles added where needed
- if two occupational titles were translated similarly, one was removed
- skill levels: indicated where needed for classification

- **The search tree**

- a 3-step search tree in the web survey for self-identification of occupation
- most easy search paths for the large & the unskilled occupations
- in each language, each tier in the tree is sorted by alphabet
- approx 1/3 of titles is placed in two or more search paths

Cut-out of WISCO database

ISCO0813	Source label	bu_BU	cs_CZ	da_DK	es_ES	fi_FI
7231030000000	Automotive glass installer or repairer	Монтажник / механик на автомобилни стъкла	Montér nebo opravář autoskel	Autoglarmester	Instalador o reparador de lunas de automóviles	Autonhuoltomies
7231010000000	Automotive body repairer	Техник-механик, автомобили	Opravář autokaroserií	Karosserieparatør	Chapista	Autopeltiseppä
7213030000000	Breaker's yard worker	Работник в автоморга	Pracovník v lomu	Ophugningsarbejder	Trabajador de desguace de automóviles	Hajottamon työntekijä
9611030000000	Vehicle spray painter	Бояджия, автомобили	Lakovač automobilů	Billakør	Barnizador de vehículos	Automaalari
7132010000000	Car seller	Дилър на коли	Prodejce automobilů	Bilforhandler	Vendedor de coches	Automyyjä
3322020000000	Bus mechanic	Автобусен техник	Autobusový mechanik	Busmekaniker	Mecánico de autobuses	Linja-autoasentaja
7231020000000	Garage manager	Началник, гараж	Vedoucí autoservisu	Garagevagt	Encargado de Taller	Korjaamotyönjohtaja
5221150000000	Diesel engine mechanic	Специалист, дизелови двигатели	Specialista na dieslové motory	Dieselmotor specialist	Especialista en motores diésel	Dieselasentaja
7231050000000	Inspector cars and other vehicles	Инспектор, коли и други превозни средства	Inspektor motorových vozidel	Inspektør af biler og andre køretøjer	Inspector de coches y otros vehículos	Auton- ja muiden kulkuneuvojen katsastaja
3359010000000	Road vehicles loader	Товарач, сухоземни превозни средства	Nakladač, závozník	Køretøjsoperatør	Cargador de vehículos de carretera	Lastaaja
9333040000000	First line supervisor mechanics, installers, or repairers	Пряк ръководител (супервайзор), механици и техници	Přímý nadřízený mechaniků, instalatérů a opravářů	Øverste tilsynsførende for mekanikere, installatører eller reparatører	Supervisor directo de mecánicos, instaladores o reparadores	Mekaanikkojen, asentajien tai korjaajien esimies

- **Critics**

- cognitive/literacy demanding & time consuming for respondent

- **Drop out rates and completion time**

- analysis 2010 Q2 WageIndicator data UK, BEL (2*), NLD (N=24,811)
- 1/3 of total drop-out, depending on number of characters read
- step 2 : median 11 seconds, mean 16, min 1, max 492
- step 3 : median 14 seconds, mean 19, min 1, max 495

- **Text box after search tree**

- 4,020 respondents entered text (22% of step-3 completions)
- 70% -> refer to composite occupations or provide detailed job descr.
- 14% -> ticked occupation is right, but occupation in text is better
- 7% -> ticked occupation is wrong (1.7% of total)
- 5% -> text not relevant for the survey question

WISCO Text string matching (TSM)

- **Reduction of drop out and response time**
 - -> single page filtering (all 3 steps in search tree on one page)
 - -> TSM (Text String Matching) -> type first letters of job title (as in Google Search), returns matches from WISCO database
 - -> respondents can chose between TSM or search tree
- **Progress in WageIndicator Salary Check**
 - TSM + search tree in Salary Check (august 2012)
 - -> few complaints from web visitors
- **Progress in WageIndicator web survey**
 - tested nov-dec 2012, launch early 2013
 - composite occupations -> multiple choices (mid 2013)

TSM in *WageIndicator-1*

1 Find your occupation.

Type the first letters of your occupation and select.

Or select your industry and occupation in the search tree.

Agriculture, nature, animals,
environment

Care, children, welfare, social
work

Cars, mechanics, technicians,
engineers

Cleaning, housekeeping,
garbage, waste

Clerks, secretaries, post,
telephone

Commercial, shop, buy and
sale

Construction, fittings, housing

Education, research, training

Finance, banking, insurance

Food manufacturing

TSM in *WageIndicator-2*

1 Find your occupation.

sec

...

- Secretary
- Security or fire alarm systems installer or repairer
- Secretary clerk
- Administrative secretary
- Executive secretary
- Legal secretary
- Medical secretary or receptionist
- Minutes secretary
- Project secretary
- Secretary, all other
- Team or department secretary
- Post-secondary education teacher agriculture, nature, animals, environment sciences
- Post-secondary education teacher computer sciences and mathematics
- Post-secondary education teacher economics, business and

1 Find your occupation.

secr

...

- Secretary
- Secretary clerk
- Administrative secretary
- Executive secretary
- Legal secretary
- Medical secretary or receptionist
- Minutes secretary
- Project secretary
- Secretary, all other
- Team or department secretary

TSM in *WageIndicator-3*

1 Find your occupation.

About this Salary Check



Or select your industry and occupation in the search tree.

Agriculture, nature, animals,
environment

Care, children, welfare, social
work

Cars, mechanics,
technicians, engineers

Cleaning, housekeeping,
garbage, waste

Clerks, secretaries, post,
telephone

Commercial, shop, buy and
sale

Construction, fittings, housing

Education, research, training

Finance, banking, insurance

Food manufacturing

Cars

Drafter

Electrical

Engineer

Engineering technician

Mechanic, electrician

Petrol station

Research, consultancy

Support services (internal)

Car wash attendant

First line supervisor automotive
parts or fuel

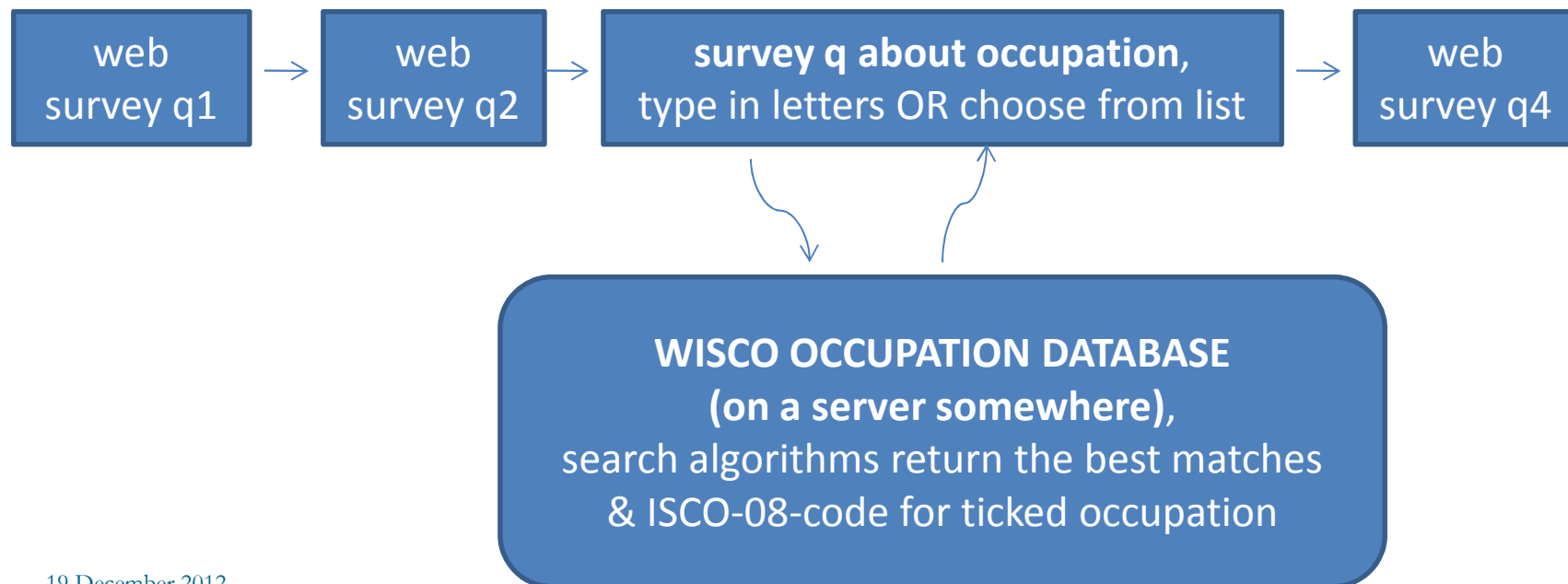
Petrol pump attendant

Sales assistant automotive
parts or fuel

Shop keeper automotive parts
or fuel

Internet tool for web surveys

- **Occupation database in the cloud (dec.12)**
 - Application Programming Interface (API)
 - any web survey can call upon this tool for its occupation question
-> ISCO-08 dgt code
 - MySQL database with search algorithms for TSM
 - why: to reduce testing time & allow for multiple choices





WISCO – problems to be solved

- **Occupations not listed in database**
 - rare or new occupational titles
 - -> to be solved by 'suggest new item' box (to be developed)
- **Web surveys on smart phones**
 - TSM with search tree too large for smart phones
 - -> to be solved with shorter match lists & tighter match algorithms
- **Occupation question for other respondents than job holders**
 - What is/was occupation father/mother?
 - What occupation are you studying for? (students, school pupils)
 - What occupation are you looking for? (job seekers)
 - -> WISCO database needs adaptation (higher aggregation level)



3) Empirics about occupations

Occupations cross-country similar?

- **ILO - ISCO**

- ISCO classifies occupations by skill level, but ILO does not investigate skill level empirically (undoable worldwide)
- ILO has few coding instructions & hardly staff involved in coding

- **Eurostat**

- National Statistical Offices are required to provide their LFS data with ISCO-08 (3 or 4 digit)
- Eurostat has a discussion platform and undertakes projects
- has meetings with NSO's about occupational coding

- **Statistics Netherlands**

- withdrew national occupational classification for ISCO, used CASCOT with WISCO-NL database, added own codes

Occupations cross-country similar?

- **Two viewpoints**

- most occupational titles are not similar -> stick to the black box approach
- most occupational titles are similar -> start a multi-country project about occupational coding

- **Few studies: are occupations similar?**

- studies -> mostly few occupations and few countries
- measurement problems, data collection problems
- assumptions in favor of comparability:
 - globalisation of the economy -> need to understand occupations
-> pressures towards cross-country standardization (e.g. QESH auditor)
 - global equipment suppliers -> tasks in jobs become similar

Empirical studies about occupations

- **EurOccupations project (2007-09)**

- are occupations similar across countries?
 - > survey of experts asking about task frequencies for 160 occupations * 10 tasks (N=4197) -> results inconclusive

- **WageIndicator data collection (2013)**

- survey of jobholders in WageIndicator web survey for XX countries, task lists for all 433 ISCO 4-dgt occupations available in 7 languages
- tasks list for each ticked occupation, asking jobholders (large N)
 - > how often they perform each task in the task list of their occupation
 - > which task requires the highest & which the lowest skill level
- aim-1: to identify task & skill profiles per occupation
- aim-2: to understand the mechanisms of division of labour in workplaces



4) The way ahead: an online tool for measuring and coding occupations



Multi-country surveys ...

- **... are at a cross-road:**
 - accept country-specific black box coding
- **OR**
 - start a joint project to develop an online tool for measuring and coding occupations



The online tool consist of

- **Coding index & occupation database**
 - a multi-country occupation database (ISCO-08)
 - with national coding indexes
- **Algorithms**
 - for cleaning text strings for misspellings & redundant words
 - for Text String Matching using coding index & occupation database
 - for auxiliary variable matching
- **A growing index & occupation database**
 - a tool for users' manual coding of under-threshold text strings
 - coding index will increase with each user
 - selection of these occupations into the occupation database
 - > requires translations of these occupational titles



Adding occupations to database

- **From coding index -> occupation database**
 - AIM: to make occupational titles as much as possible comparable across countries
 - the titles newly added to a national coding index will be judged for uniqueness & frequency
 - based on an ‘add new occupation’ wizard with few related questions
 - using a discussion platform of national experts
 - Eurostat’s CIRCA platform has shown that the discussions about coding an occupation is viable across countries
 - then translated & added to the occupation database
- **The quality of the occupation database**
 - each newly added occupation increases the quality of the TSM
 - improves the comparability of occupations in multi-country surveys



Coding algorithms

- **Cleaning**

- text strings are cleaned for misspelling and redundant words
- -> misspelling algorithms available for quite some languages
- -> redundant word algorithms to be developed /available?

- **Text string matching**

- text strings matched with occupational titles in WISCO database
- -> matching scores, choice of 3-5 best matching occupations

- **Auxiliary variable matching**

- for these 3-5 best matching occupations a matching score based on gender, age, ISCED, NACE, firm size, empl. status is provided

- **Manual coding for match under threshold**

- archiving the manual coding -> building coding index



The online tool serves

- **Coding of Open Ended Questions**

- text strings from Open Ended Questions in multi-country and single-country surveys
- using field- or for office coding
- returns for each respondent ISCO-08 5-dgt code to survey holder

- **Self-identification**

- an API for use in web surveys with TSM & search tree
- or to be used by interviewers in CAPI or CATI (assumes interviewer training)
- returns for each respondent ISCO-08 5-dgt code to survey holder



Example

- **An agency with a survey ...**
 - derives a new dataset with:
 - survey ID, locale (=language per country)
 - respondent ID, text strings for occupation (+/- job description)
 - gender, age, education/ISCED, NACE2.0, firm size, empl. status
- **... uploads this file in the coding tool**
 - the survey agency codes manually the under-threshold cases using suggestions from the tool
 - the tool returns a file with respondent ID + ISCO-08 5 digit codes
- **.. and is meanwhile contributing**
 - this agency profits from the occupation database and from coding efforts of previous users of the tool, but contributes its coding efforts for future users (conditional for use of the tool)



Organisational framework

- **Joint efforts are a primary condition**
 - joint activity of many multi-country surveys (ESS, PIAAC, PISA, WVS, EVS, EWCS, SHARE, ISSP, others)
 - aiming for cooperation with Eurostat (with CIRCA platform), CEDEFOP, ESCO, EuroFound, others
 - pooling the expertise of CASCOT, WageIndicator, others
- **Organisation**
 - long term commitment (possibly to the ISCO 2028 update)
 - online platform must be permanent available
 - hosted by well-known institution
 - network organisation with occupation expert in each country
- **Business model to be developed**



Other services of the tool

- **Cross-over tables for occupation**
 - eg social stratification
 - gender or age composition tables, etc
- **Events, studies, documentation, etc**
 - draft design for empirical studies on occupations (freq tabels across countries, educational attainments at ISCO-4 dgt,.....)
mapping occupational composition and structures across countries
- **Network**
 - network of occupational coding experts & discussion platform
- **A similar tool to code industries**
 - the tool could be expanded for coding NACE or ISIC



Problems to be solved

- **Composite occupations**
 - Composite job titles -> 1+1 title to be ticked?
 - Division of labour across job titles depends on firm size
- **Managerial & self-employed occupations**
 - Management job titles & other large categories
 - Job titles of self-employed -> ownership merges with job title
- **Coding of occupations of non-jobholders**
 - Respondent's interest -> does a job title make sense to all?
 - Job titles of unemployed -> Q: For which occ's are you educated?
 - Job titles of fathers / mothers / job-seekers / students, etc.



Thank you for your attention

- **Is this online tool for measuring and coding occupations needed and within reach?**
 - Comments invited!
- k.g.tijdens@uva.nl
- For more information [www.wageindicator.org](http://www.wageindicator.org/main/researchlab/occupation-data-base)
www.wageindicator.org/main/researchlab/occupation-data-base