# Self-identification of occupation in web-surveys: Respondents' choice between autosuggest and search tree

Amsterdam Institute for Advanced labour Studies
University of Amsterdam

*2017 CSDI Workshop Mannheim*

*Kea Tijdens,*
*University of Amsterdam/AIAS*

*17 – 03 - 2017*

# Setting the scene

- **The national stocks of job titles are …**
  - large >> 10,000's of job titles in any national labour force
  - unstructured >> vague boundaries between job titles
  - unlimited >> no fixed list, many entries and exits over time
- **The challenge for multi-country surveys**
  - to classify job titles into ISCO-08 classification of occupations
  - & ... to do so consistently across countries
- **Occupational titles vs job titles**
  - job titles: within organisational context
  - occupational titles: beyond organisational context

# ‘What is your occupation?’

- **Open-ended survey questions: textbox**
  - Textbox: used in most surveys
  - Office coding needed: expensive and time-consuming, though increasingly high quality coding software and auto-coders
  - Coding problems: vague job titles, highly aggregated titles, company-specific titles & typos
- **Closed survey questions: dictionaries**
  - Brief list (max 10 entries): predominantly used in postal surveys >> aggregation bias
  - Showcard (max 50 entries): predominantly used in face-to-face surveys
  - Dictionary (unlimited number of entries): respondents self-select their occupation from a list of occupations

# The WageIndicator web survey in 91 countries

## WageIndicator websites

- In 2001 in NL: website with job related content, salaries per occ.
- Today: web portal with national websites in 91 countries, all in national language(s)
- 2016: 40 million of visitors, most through search engines

## Multilingual, continuous web survey

- All websites invite visitors to complete (>> opt-in survey)
  - a long salary survey with lottery incentive
  - or a mini-survey to get a salary indication
- Large numbers of survey respondents
- Over the past 15 years tremendous technological progress
- Survey frequently used for micro-targeting specific respondents through social media and through face-to-face surveys with app

# Closed survey question: "What is your occupation?"

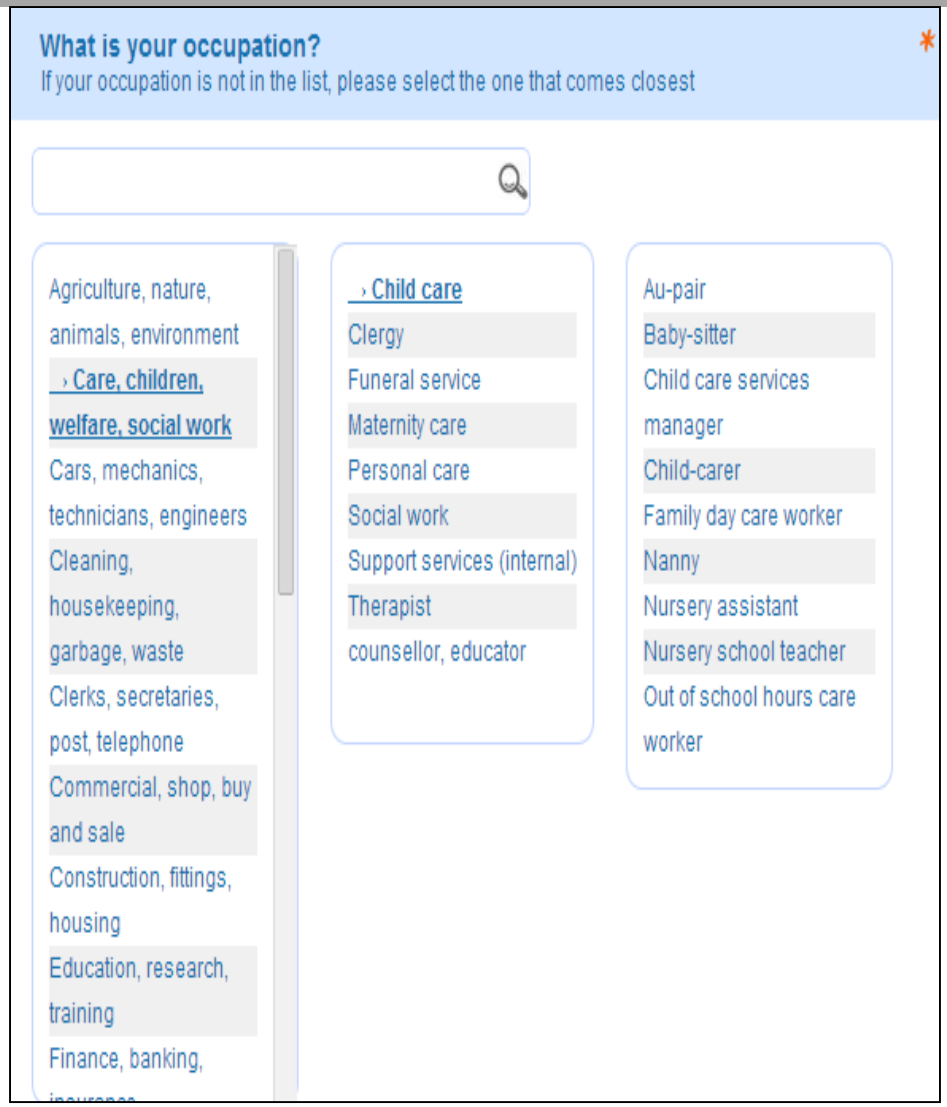
## Self-identification: search tree or autosuggest

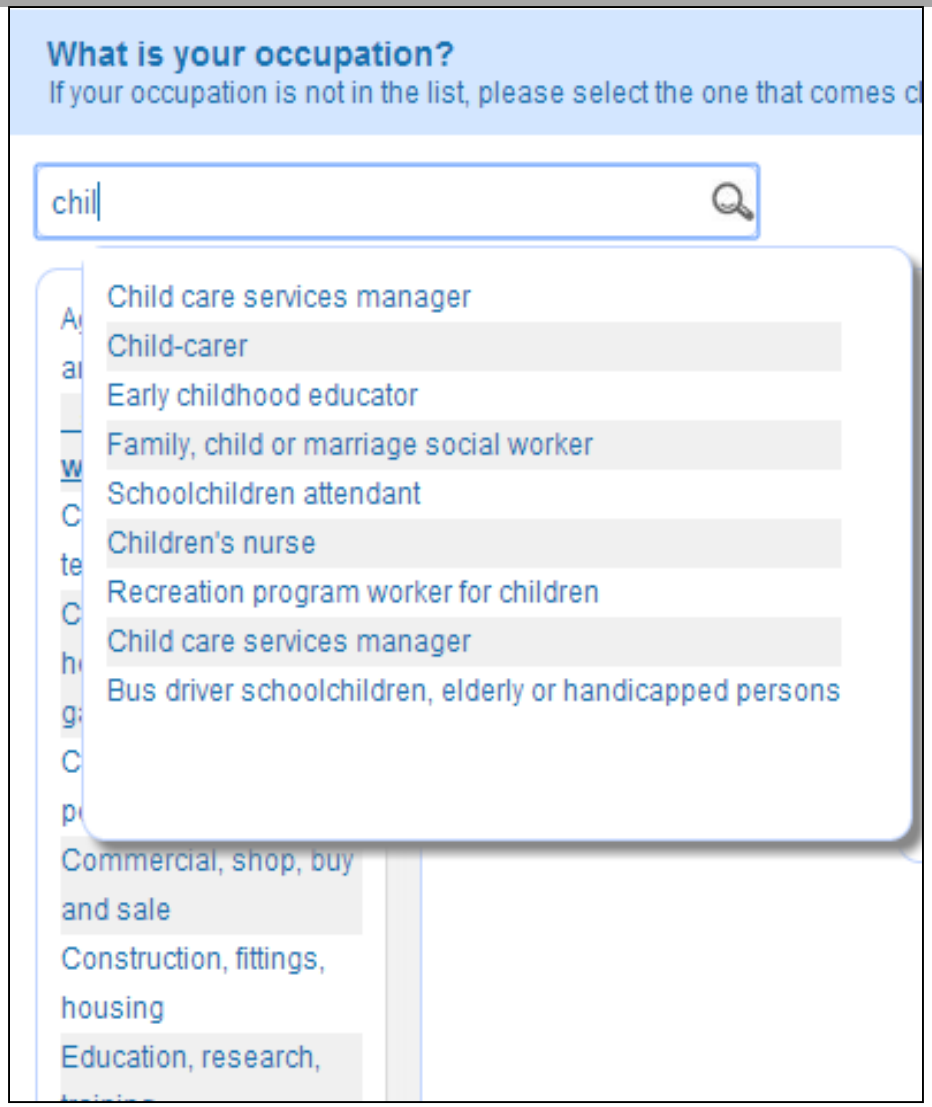
- Closed survey question, because coding was too expensive
- Respondents self-identify their occupation through
  - a search tree (IPod menu) or
  - an autosuggest box (Google search type)
- Dictionary: multi-lingual database of occupations coded ISCO08

## • History of the closed question & database

| Years | # occ's | # lang. | search tree | auto s. | API |
|---|---|---|---|---|---|
| 2001-'05 | 700 | 1 | 2-level, 1 page p/lvl | - | - |
| 2006-'09 | 1,100 | 8 | 3-levels, 1 page p/lvl | - | - |
| 2009-'12 | 1,600 | 30 | 3-levels, 1 page p/tree | - | - |
| 2012-'15 | 1,600 | 30 | 3-levels, 1 page p/tree | yes | - |
| 2015-'17 | 1,600 | 43 | 3-leves, 1 page p/tree | yes | yes |

# Search tree (left), autosuggest (right)



# Autosuggest in Chinese …



# Search tree vs autosuggest

- **Meta Data occupation API in WageIndicator web survey**
  - API data 26 June – 3 Nov 2016, selection nl_NL
  - API data registers each click and each character in autosuggest
  - 18 448 records of 2994 respondents
- **Descriptives users**
  - One in four uses a mobile device (25.6%)
  - Two in three start using the search tree (66.9%)
  - Mobile users use less often search tree (Chisq 57.37, sign .000)

| | No mobile | Mobile | Total |
|---|---|---|---|
| Search tree | 70.7% | 55.8% | 66.9% |
| Autosuggest | 29.3% | 44.2% | 33.1% |
| Total | 100.0% | 100.0% | 100.0% |

# Clicks in the search tree

- **Do respondents go back and forth in the search tree?**
  - 2004 respondents started the search tree
  - 3% dropped out with one or two clicks
  - 54% found their occupation in three clicks
  - 15% went back and forth one time
  - 28% went back and forth more than one time

| clicks | 0 | b&f1 | b&f>1 | Total |
|---|---|---|---|---|
| 1 | 31 | 0 | 0 | 2% |
| 2 | 28 | 0 | 0 | 1% |
| 3 | 1081 | 0 | 0 | 54% |
| 4 | 0 | 301 | 0 | 15% |
| 5 | 0 | 0 | 174 | 9% |
| 6 | 0 | 0 | 94 | 5% |
| 7 | 0 | 0 | 77 | 4% |
| 8 | 0 | 0 | 40 | 2% |
| 9 | 0 | 0 | 26 | 1% |
| 10 | 0 | 0 | 152 | 6% |

# Response time

- **Response time in seconds**
  - **After controlling for drop outs and outliers (<1 - >360 sec.), response times of 2817 respondents were analysed**
  - **Mean response time equal for autosugggest and search tree (39 sec.)**
  - **Median response time larger for search tree than for autosugggest (24 versus 16 sec.)**

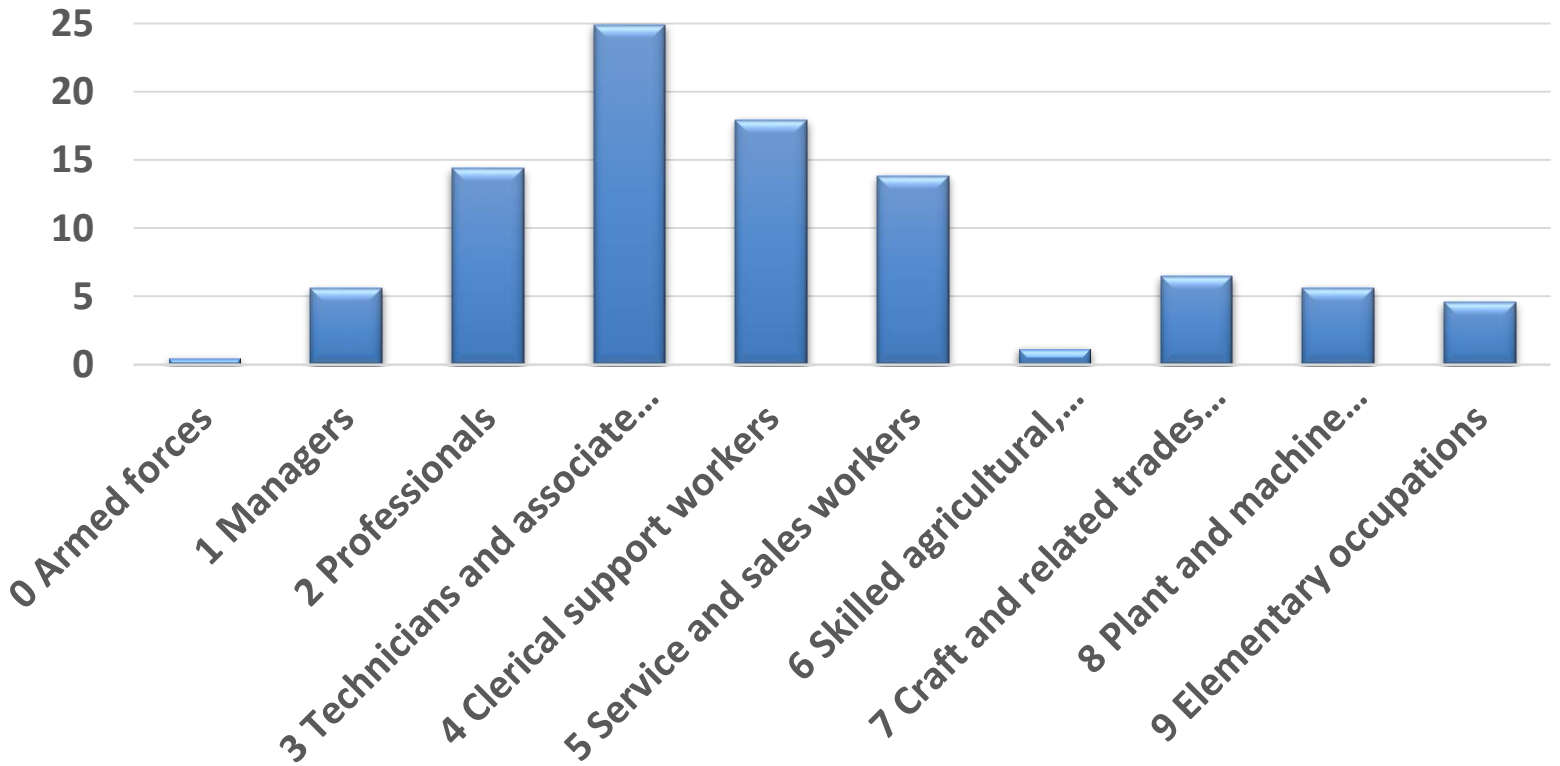
| | Minimum | Median | Mean | Maximum |
|---|---|---|---|---|
| Search tree | 1.7 | 23.6 | 39.1 | 347.0 |
| Autosuggest | 1.1 | 16.4 | 39.1 | 335.9 |
| Total | 1.1 | 21.6 | 39.1 | 347.0 |

# Drop out

- **Which respondents drop out?**
  - 5% of respondents drop out when they self-identify their occupation
  - No sign difference between mobile and non-mobile users
  - No sign difference between search tree and autosuggest users
  - Search tree users are more likely to drop out after five clicks
  - Autosuggest users are more likely to drop out after typing 20 characters

# Selected occupations

- **Selected occupations ISCO-08**
  - The 2847 respondents selected 724 unique titles from the list of 1,600 titles (ISCO 5 digit)
  - Graph shows the distribution (%) for 1-digit ISCO-08



# SERISS project (2015-2019)

- **Extend the dictionary of occupations**
  - To 99 countries with 47 languages
  - To serve self-identication through search tree and autosuggest
  - Ensure that all occupational titles are well coded in ISCO-08
- **Make database available for survey holders**
  - Program API for web surveys on desktop, tablet, smartphone
  - Program an interface for use in CAPI surveys
  - Make database downloadable in excel
  - Availability: till end SERISS free of charge
- **Develop an occupation – industry prediction**
  - Depending on ticked occupation, a limited set of industries is shown for the survey question ‘In which industry do you work?’
  - Aiming to reduce respondents’ time

# SERISS – databases: industry, education, employment status

- **What is main activity of organisation where you work?**
  - Industry database with 320 entries
  - Coded 3 or 4 digit NACE2.0, crossovers to ISIC
  - Database & related survey question available in 47 languages
- **What is your highest education?**
  - >> See presentation of CAMCES database
- **What is your employment status?**
  - Survey questions & answers to identify employment status
  - Coded ICSE-93 (ILO classification) + prospectieve update
  - Coded ESEG-2014: Eurostat's socio-economic classification (jointly with occupation code)
  - Survey questions & answers available in 47 languages

# Available for survey holders?

- **Yes, survey module**
  - The full module consists of 32 survey questions
  - This includes questions for multi-country surveys: selection of language and country of residence
  - This includes questions for routing: year of birth, gender, in paid employment/unemployed
  - This includes questions in present and past tense for current or latest job
  - Survey holders can select questions they want to ask
  - SERISS provides free of charge until Jun2019
- **Workshop 4-5 September in Amsterdam**
  - Presentation of the module and demo
  - Presentations of databases
  - SERISS can provide travel&accomodation funding

# The end .....

## Thank you for your attention ☺ ☺

## Useful links

http://www.wageindicator.org/main

## Please try the demo:

- **http://tmt.centerdata.nl/jobcoder_demo/**

## Questions?


- **k.g.tijdens@uva.nl**