WP8 An occupation database for multi-country surveys

Kea Tijdens, UVA / WageIndicator
Michele Belloni, University Ca' Foscari of Venice
SERISS WP8 Workshop in Venice
14 February 2019
Outline

• What is an occupation?
  – Why measuring occupations in surveys?
  – Classification & harmonization across countries

• Contribution of surveycodings.org / SERISS

• Concepts
  – history of the occupation database
  – source list & translations of the source list
  – occupation – industry prediction

• How surveyholders can use the database?
  – Browse to Database live search and select country
  – Use autocomplete box & Search tree
  – Result in dataset
What is an occupation?

• Definition
  – A job title comprises of a bundle of tasks, in context of the workplace
  – An occupation identifies a bundle of job titles, distinct from another occupation
  – Occupational titles are used for communication beyond the workplace

• The national stocks of occupations are ...
  – large >> 10,000’s of occupations in any national labour force
  – unlimited >> no fixed list, many entries and exits over time
  – unstructured >> vague demarcation lines between job titles
  – distribution >> very skewed: many nurses, few C+ programmers
What is an occupation – multi-country

• The challenge
  – to classify occupations into the 433 occupational units in the ISCO-08 international classification of occupations & ... to do so consistently across countries

• Problem: occupation demarcation lines vary across countries
  – Educational system (contribution of vocational versus generic systems)
  – Collective bargaining systems (pay scales according to job titles)
  – Most common firm sizes (large firms have more division of labour vs small firms)
  – Licensing systems (varies largely in EU)

• But .. occupation tend to converge across countries
  – Increasing global suppliers standardizing the related occupations
  – Increasing pressure for standardisation (sea / air transport)
  – Increasing global communication (IT, tech, hotels)
  – Increasing use of English job titles (e.g. rapper, singer-songwriter)
Why measuring occupations in surveys?

• Important concept in social sciences
  – For identifying social status, understanding wages, vocational systems, etc

• Answering the “What is your occupation” question
  – Open-ended: textbox with field or office coding
    • expensive, time-consuming
    • for single country surveys increasingly high quality coding software and auto-coders
    • coding problems: 1-10% of observations is unidentifiable; 1-10% is coded at higher aggregation level
  – Closed: dictionary with occupational titles (coded)
    • max 10 entries in postal surveys, max 50 entries in showcard, unlimited in web surveys

• Answering varies by target group
  – Working population > answering own occupation in detailed categories
  – Not working population > answering last/preferred occupation, education
  – Occupation of father/mother/partner > answering aggregated categories
Classification & harmonization across countries

• The international classification ISCO
  – Maintained by ILO: International Standard Classification of Occupations (ISCO)
  – Updates every 20 years: ISCO-68, ISCO-88, ISCO-08, ISCO-28 (scheduled)
  – ISCO-08: Coding index in English, no multilingual index available
  – ILO: No ISCO support staff or discussion platform (poor compared to NACE class.)

• Coding practices in multi-country surveys
  – Multi-country coding indexes do not exist
  – SERISS D8.4: 99 countries >> only 34 publicly available national coding index >> of which 19 indexes 5-digit ISCO-08 >> of which 16 more than 1000 entries >> these 16 indexes coded largely different sets of job titles
  – ISCO coding is the standard in multi-country surveys
  – Some countries use their national classification with national coding index (USA, UK, DE, FR, IT, PL)
Classification & harmonization across countries

• Coding practices in multi-country surveys: office coding
  – Most surveys use open-text box
  – National survey agencies perform own office coding into ISCO-08
  – >> black box: are the same occupations coded similarly across countries?

• Validating through translation of occupations: two viewpoints
  – Occupational titles cannot be translated: countries have different occupational titles beyond ISCO-4-digit >> the black box cannot be solved
  – Occupational titles can be translated as refer to similar set of activities >> the black box can be validated
  – Examples of validating exercises
    • Eurofound EWCS survey 2010 for associate EU countries translating titles into English
    • ESS father/mother occupations (Ganzeboom)
  – SERISS D8.3: multilingual, multi-country coding index of occupational titles
Contribution of surveycodings.org / SERISS

- **Multi-country occupation database**
  - Build a coding index of occupational titles, coded ISCO-08, for self-identification
  - Departing from ISCO-08 classification
  - Add occupational titles from non-English indexes
  - Translate titles in all languages (99 countries, 47 languages)
  - For most languages 4 000+ titles

- **Multi-country industry database**
  - Build a coding index of industries, coded NACE 2.0, for self-identification
  - Departing from NACE industry classification
  - Prepare database of 3- or 4-digit industries (321 entries)
  - Collect translations and add missing languages (99 countries, 47 languages),

- **Use the databases for self-/interviewer-identification**
  - For CAPI surveys: use databases to implement search trees or autosuggest in CAPI module for respondent self-identification or for interviewer identification
  - For CAWI surveys: implement the API’s of the databases in web-survey for self-identification
  - Live-search demo tool is provided at surveycodings.org
Concepts – history of the occupation database

• **Start early 2000s**
  – Wageindicator Foundation runs a continuous, volunteer websurvey on work and wages, asking ‘What is your occupation’ and respondents self-select their occupation from a dictionary
  – First survey targeted working women in NLD with list of 100 occupational titles and search tree
  – Survey expanding to 8 EU countries and database to approx 1,000 titles, coded ISCO-88, and translated by national labour market experts

• **The 2010s**
  – Wageindicator websurvey expanded to 90 countries with 35 languages, database to 1,700 titles, coded ISCO-08, search tree plus autosuggest box (Google search type), since 2016 database in API
  – Since early 2000’s approx 5 mln web visitors have self-identified their occupation using Salary Survey plus Salary Check, few complaints ‘my occupation in not in your database’

• **Expanding the occupation database**
  – InGRID project (EU FP7 2014-’16) : funded further development of the occupation database design
  – SERISS project (EU H2020 2016-’19) : expanding database to 5,000 titles and 99 countries
  – SSHOC project (EU H2020 2019-’22) : expanding database for measuring parent’s occupations (20th century occupational titles)
Concepts – the source list

• **Source list**
  – Extending the source list from 1,700 to 5,000 occupations
  – Using titles from ISCO-08, national coding indexes
  – Compare codes across indexes
  – Remove duplicate/overlapping titles >> source list 4,233 entries
  – Check inconsistent codes with partner IER U Warwick

• **Principles underlying the source list (see D8.3)**
  – Skill-level differences by using different wording
  – Supervisory positions (managers vs first-line supervisors)
  – Coding of emerging occupations (blogger)
  – Level of detail for machine-operators, shop-assistants
Concepts – translations of the source list

- **Translating the source list**
  - Depart from translations available for 1,700 titles (35 languages)
  - Translate 1,700 + 2,300 titles by professional translation agency OHT (OneHourTranslation)
  - Check 1,700 titles and translate 2,300 titles for 5 main non-EU languages (IER U Warwick)

- **Comments on translating occupational titles**
  - Translating occupational titles assumes knowledge of the job content, but across countries empirical data about job content is missing
  - Some occupational titles are country–specific and therefore difficult to translate
  - Two source titles may be translated similarly (accountant vs bookkeeper), whereas other titles are unknown and therefore cannot be translated (regional police officer)
  - Translations were checked and duplicates removed
Concepts – the occupation – industry prediction

• The industry database
  – Database for survey question about ‘main business activity of the organisation where you work’, coded NACE 2.0
  – Search tree, no autosuggest (respondents report company name)
  – Difficult question for respondents, in web surveys they might skip question

• Predicting industry from occupation
  – To reduce the options for the self-selection of industry, we developed an occupation-to-industry prediction
  – We pooled datasets from EU countries, resulting in 2 mln obs with data on occupation and industry (ISCO-08 4-digit and NACE 2.0 2 digit)
  – For each 4-digit occupation, the most likely industries could be computed
  – Given their selected occupation, respondents can now chose from a limited set of industries
  – The list has an option ‘other’, leading to the full industry database
  – The prediction not yet available in the livesearch tool
How surveyholders can use the database?

• **Website surveycodings.org**
  – Read the explanatory text
  – Download the excel for own use
  – Use the live search
  – Use the API in your web-survey
  – Discuss possibilities to use the databases in CAPI software

• **What the database cannot do**
  – No mapping table to national occupational classifications
    (e.g. France, Germany, UK, Poland)
  – The excel file can be used for office coding, but no script for removing
typing errors & unidentified titles need to be coded manually
  – IER U Warwick offers its CASCOT tool for office coding, including
    several languages
Browse to Database live search and select country

Occupation

Database live search

Please select a country/language

Argentina

Autocomplete

Buscar

Tree

- Agricultura, naturaleza, animales, medio ambiente
- Asistencia, niños, protección social, trabajo social
- Coches, mecánicos, técnicos, ingenieros
Use autocomplete box

Autocomplete

hort|

Gerente de producción hortícola
Científico horticultor
Técnico de horticultura
Productor hortícola
Trabajador de vivero hortícola
Horticultor (agricultura de subsistencia)
Search tree

Tree

- Agricultura, naturaleza, animales, medio ambiente
  - Servicios de apoyo (interno)
  - Ganado
    - Horticultura, jardinería
      - Arquitecto de paisajes
      - Arquitecto de jardines
      - Florista
      - Técnico de horticultura
      - Productor hortícola
      - Jardinero ornamental
      - Jardinero de cementerio
Result in dataset

Information

Name: 6310000100018
Description: Vegetable farmer (subsistence farming)
Translated description: Horticultor (agricultura de subsistencia)
Classifications:
  - ISCO08: 6310 - Subsistence crop farmers

View coding and related variables in L1VE database
Questions??

See https://www.surveycodings.org/

Thank you for your attention

Kea Tijdens k.g.tijdens@uva.nl