Developments in the (Semi-)automatic coding of occupations

Eric Harrison
on behalf of DASISH WP3 Team
City University London/CentERdata/IER Warwick

CSDI 2015
London
What are the problems with occupation coding?

- Occupation is a standard measure on all social surveys
- Important in its own right and as a building block for social class measures
- Complicated to collect and in non-standard form
- Requires harmonisation to (max) four-digit classification
- Requires specialist knowledge to code accurately
The DASISH Project

- Three year FP7 cluster project for SS&H
- All work packages aiming to focus on common ground and pool expertise
- Social science WP on Survey Quality: producing software to improve survey management, data collection and user documentation
- Ultimate goal – ‘interoperability’
- 3.1. Software for improved coding of occupation (ESS, SHARE, CESSDA, Warwick)
3.1. Occupation Coding: original brief

- To develop software for Europe-wide surveys to collect and code occupation data more accurately, consistently and cost-effectively
- Key was to enable better coding at time of collection
- Two directions in original bid:
  - CAPI tool allowing interviewer to select best match to text from respondent
  - Web survey based occupation tree that was trialled in the Eurooccupations project
- Open source code for use in different CAPI languages
Strands of work

- Colleagues at Warwick working on upgrade of existing CASCOT software for use internationally (v 5.0)
- Expert ‘volunteers’ in selected countries assisting with translating terms, classification files and testing software
- SHARE team in Venice checking ISCO coding against CASCOT
- Colleagues at Tilburg pursuing related work on text matching and apps for handheld devices
- Wider network of colleagues via InGRID project – database of job titles, skills and tasks
Developing ‘CASCOT International’

- A new facility within CASCOT: 
  - to detect automatically and switch the interface language 
  - to handle various language classification files

- The international version of CASCOT was supplied to and evaluated by national occupational experts in relevant countries
‘CASCOT International’

- Job titles in the selected languages indexed to ISCO 08
  - Some supplied by NSIs or other groups
  - Some contacted directly through website information
- Raw data files from the European Social Survey (ESS) Round 6 were supplied to Warwick. Some contain job titles and ISCO 08 codes. Used to validate the software.
- Compare manual and CASCOT codes – look at ‘performance’ and ‘quality’
- Development of performance tool allows users to generate charts

www.europeansocialsurvey.org
CASCOT development in DASISH

- User interface in 8 languages:
  - Dutch, English, Finnish, French, German, Italian, Slovak and Spanish

- ISCO-08 classification (structure, index) prepared for each country

- Simultaneous coding into ISCO-08 and national code possible
CASCOT development in DASISH

- Development of CASCOT Performance Tool
- Raw data files from the European Social Survey (ESS) Round 6 used to validate the software
- Partnership arrangements for the testing and fine-tuning by experts within each country covered by the languages in the pilot
Selecting interface language

Then restart CASCOT
Selecting classification

Select from the menu ‘Classification’ and choose from the list. If the desired classification is not listed, select File>Open classification, navigate to the correct folder, select the desired classification file and click ‘Open’. 
Selecting output items

Select Options>Output
And click ‘Add’ next to the items you wish to have in the output.

NB National code can be added to the output as in this example.
Current output is shown at the bottom, click ‘Ok’ to accept.
Parallel work: CentERdata’s Jobcoder app

- Opting for text matching rather than decision tree
- App - Easily integrated in web and capi interview
- Can also help with paper-based interviews
- No need to be online
- Low cost

www.europeansocialsurvey.org
Venice event 2014

- Key stage in improvement of international coding standards = expert input and databases of titles
- Workshop in Venice to discuss and fine-tune CASCOT
- Presentations show ceiling of 70% for matching to 1 digit ISCO
- Hands on fine tuning session focused on creation/identification of new coding rules in range of languages
Jobcoder for CAPI

ISCO Coder

Far

Farmer (mixed farming)
Farm machinery repairer
Farm labourer
Cereal farmer
Cotton farmer
Rice farmer
Skilled farm worker (field crops)
Fruit farmer

Fruit farmer

Fruit farmer (6112)

6: Skilled agricultural, forestry and fishery workers

61: Market-oriented skilled agricultural workers

611: Market gardeners and crop growers
Jobcoder embedded in CAWI
Future

- Horizon 2020 call – ‘Infradev4’
- ‘Synergies for European Research in the Social Sciences’ (SERISS)
- This time exclusively social science based: ESS, SHARE, CESSDA, GGP, EVS – and UVA
- Wide range of issues relating to survey: sampling, translation, data collection, legal-ethical, plus a WP to create a coding module and supporting databases – for occupation but also broader labour market info
Look forward to reporting back!

Eric.harrison.2@city.ac.uk
www.europeansocialsurvey.org