

# **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

# CASCOT development in DASISH

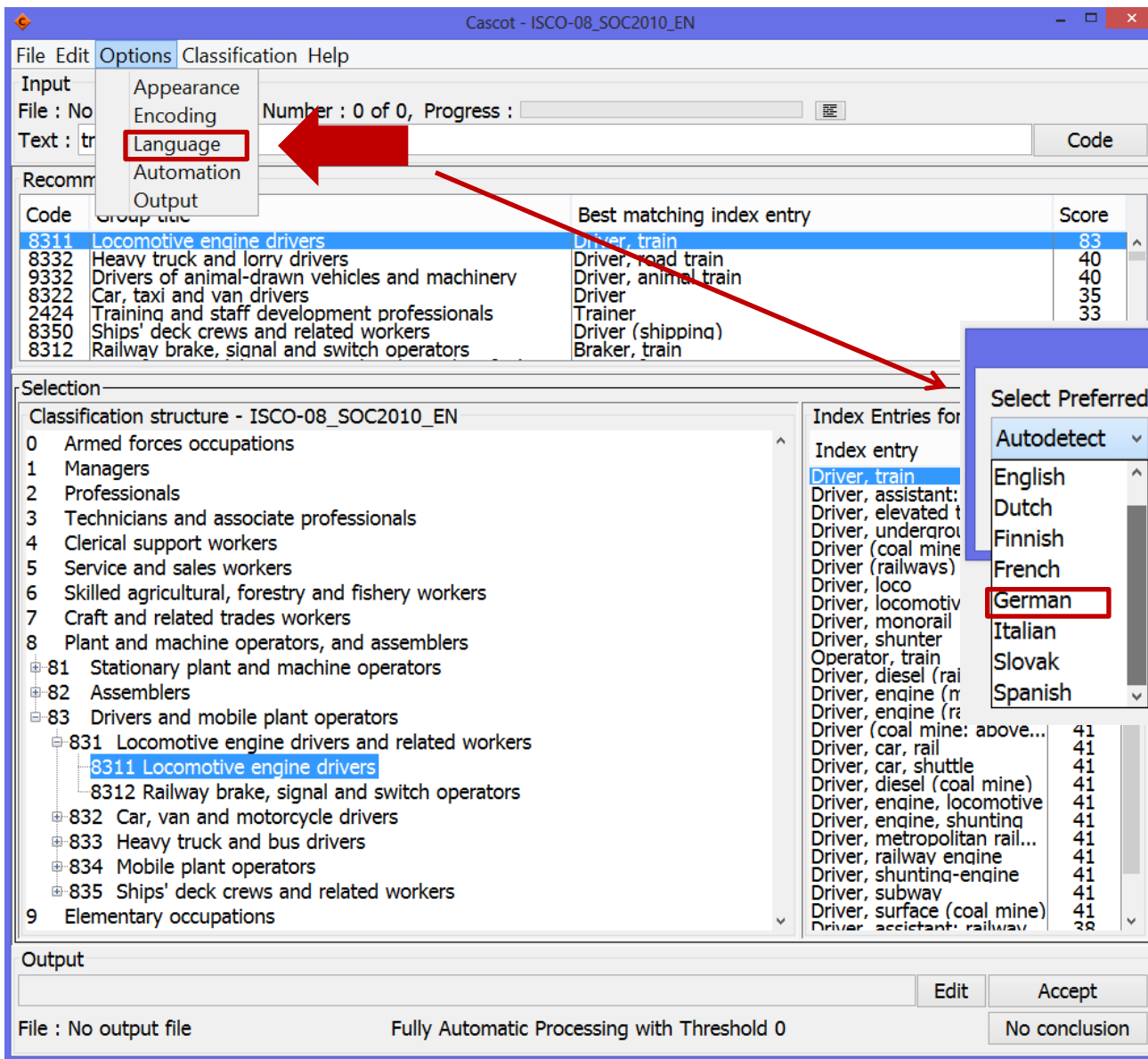
- **User interface in 8 languages:**
- Dutch, English, Finnish, French, German, Italian, Slovak and Spanish
- **ISCO-08 classification (structure, index)** prepared for each country
- Simultaneous coding into ISCO-08 and **national code** possible



# 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



The screenshot shows the Cascot software interface. The 'Options' menu is open, and 'Language' is selected. A red arrow points from the 'Language' option in the menu to the 'Language' dialog box. In the dialog box, 'German' is selected in the 'Select Preferred Language' list. Another red arrow points from the 'Language' dialog box to the 'Output' field at the bottom of the main window.

**Options Menu:**

- Input
- File : No
- Text : tr
- Recomm
- Appearance
- Encoding
- Language**
- Automation
- Output

**Language Dialog Box:**

Select Preferred Language

Autodetect

English

Dutch

Finnish

French

**German**

Italian

Slovak

Spanish

Cancel

**Main Window:**

Cascot - ISCO-08\_SOC2010\_EN

File Edit Options Classification Help

Number : 0 of 0, Progress : [Progress Bar]

Code	Group title	Best matching index entry	Score
8311	Locomotive engine drivers	Driver, train	83
8332	Heavy truck and lorry drivers	Driver, road train	40
9332	Drivers of animal-drawn vehicles and machinery	Driver, animal train	40
8322	Car, taxi and van drivers	Driver	35
2424	Training and staff development professionals	Trainer	33
8350	Ships' deck crews and related workers	Driver (shipping)	
8312	Railway brake, signal and switch operators	Braker, train	

**Selection:**

Classification structure - ISCO-08\_SOC2010\_EN

- 0 Armed forces occupations
- 1 Managers
- 2 Professionals
- 3 Technicians and associate professionals
- 4 Clerical support workers
- 5 Service and sales workers
- 6 Skilled agricultural, forestry and fishery workers
- 7 Craft and related trades workers
- 8 Plant and machine operators, and assemblers
  - 81 Stationary plant and machine operators
  - 82 Assemblers
  - 83 Drivers and mobile plant operators
    - 831 Locomotive engine drivers and related workers
      - 8311 Locomotive engine drivers**
      - 8312 Railway brake, signal and switch operators
    - 832 Car, van and motorcycle drivers
    - 833 Heavy truck and bus drivers
    - 834 Mobile plant operators
    - 835 Ships' deck crews and related workers
- 9 Elementary occupations

**Index Entries for:**

Index entry	Score
Driver, train	83
Driver, assistant:	
Driver, elevated t	
Driver, underground	
Driver (coal mine	
Driver (railways)	
Driver, loco	
Driver, locomotiv	
Driver, monorail	
Driver, shunter	
Operator, train	
Driver, diesel (rai	
Driver, engine (r	
Driver, engine (r	
Driver (coal mine: above...	41
Driver, car, rail	41
Driver, car, shuttle	41
Driver, diesel (coal mine)	41
Driver, engine, locomotive	41
Driver, engine, shunting	41
Driver, metropolitan rail...	41
Driver, railway engine	41
Driver, shunting-engine	41
Driver, subway	41
Driver, surface (coal mine)	41
Driver, assistant: railway	38

**Output:**

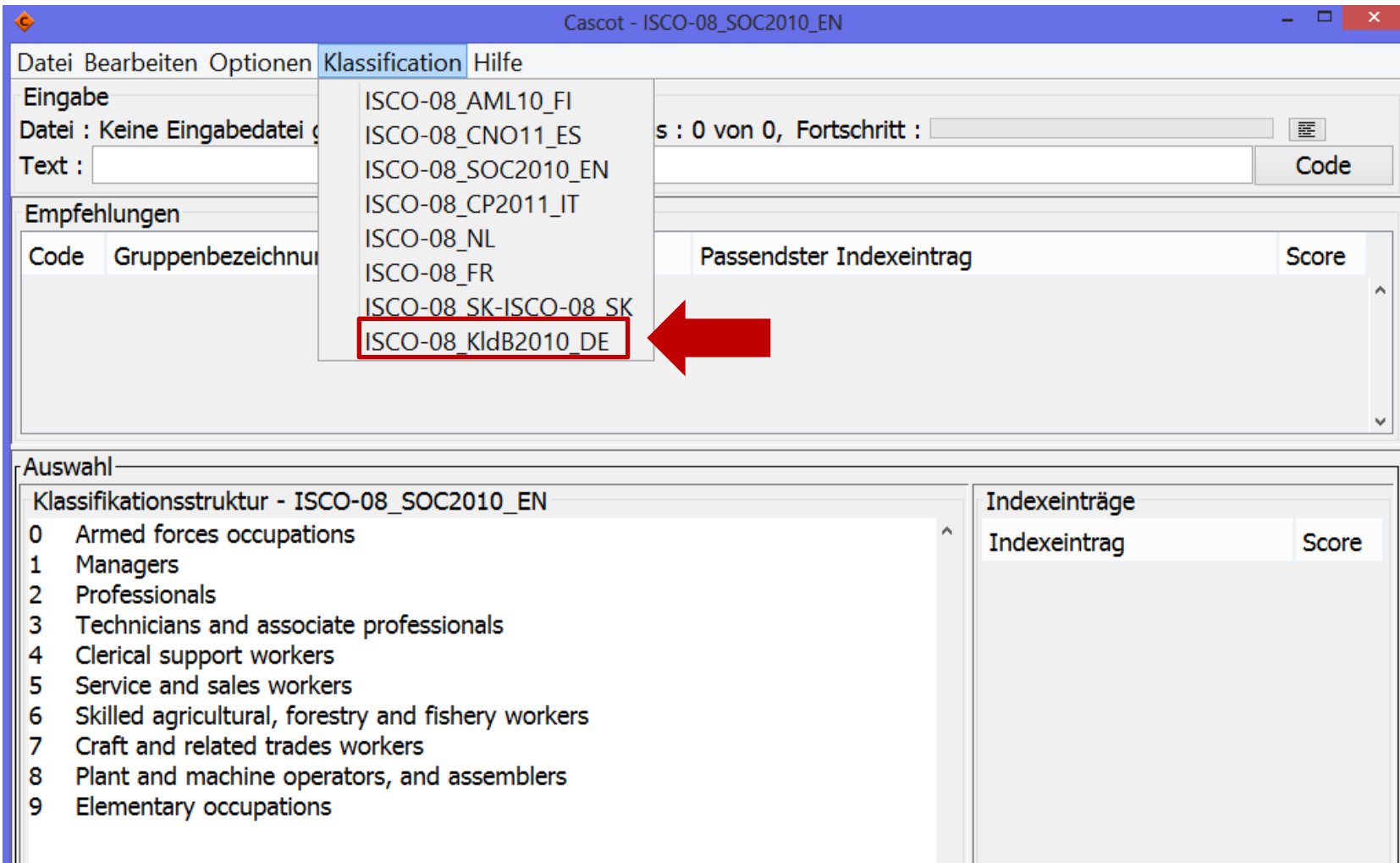
File : No output file Fully Automatic Processing with Threshold 0

Edit Accept

No conclusion

Then restart  
CASCOT

# Selecting classification



The screenshot shows the 'Cascot - ISCO-08\_SOC2010\_EN' application window. The 'Klassifikation' menu is open, displaying a list of classification files. A red box highlights 'ISCO-08\_KldB2010\_DE', with a red arrow pointing to it from the right. Below the menu, the 'Auswahl' panel is visible, showing a tree view of the 'Klassifikationsstruktur - ISCO-08\_SOC2010\_EN' and an empty 'Indexeinträge' table.

**Klassifikation Menu:**

- ISCO-08\_AML10\_FI
- ISCO-08\_CNO11\_ES
- ISCO-08\_SOC2010\_EN
- ISCO-08\_CP2011\_IT
- ISCO-08\_NL
- ISCO-08\_FR
- ISCO-08\_SK-ISCO-08\_SK
- ISCO-08\_KldB2010\_DE**

**Auswahl Panel:**

Klassifikationsstruktur - ISCO-08\_SOC2010\_EN

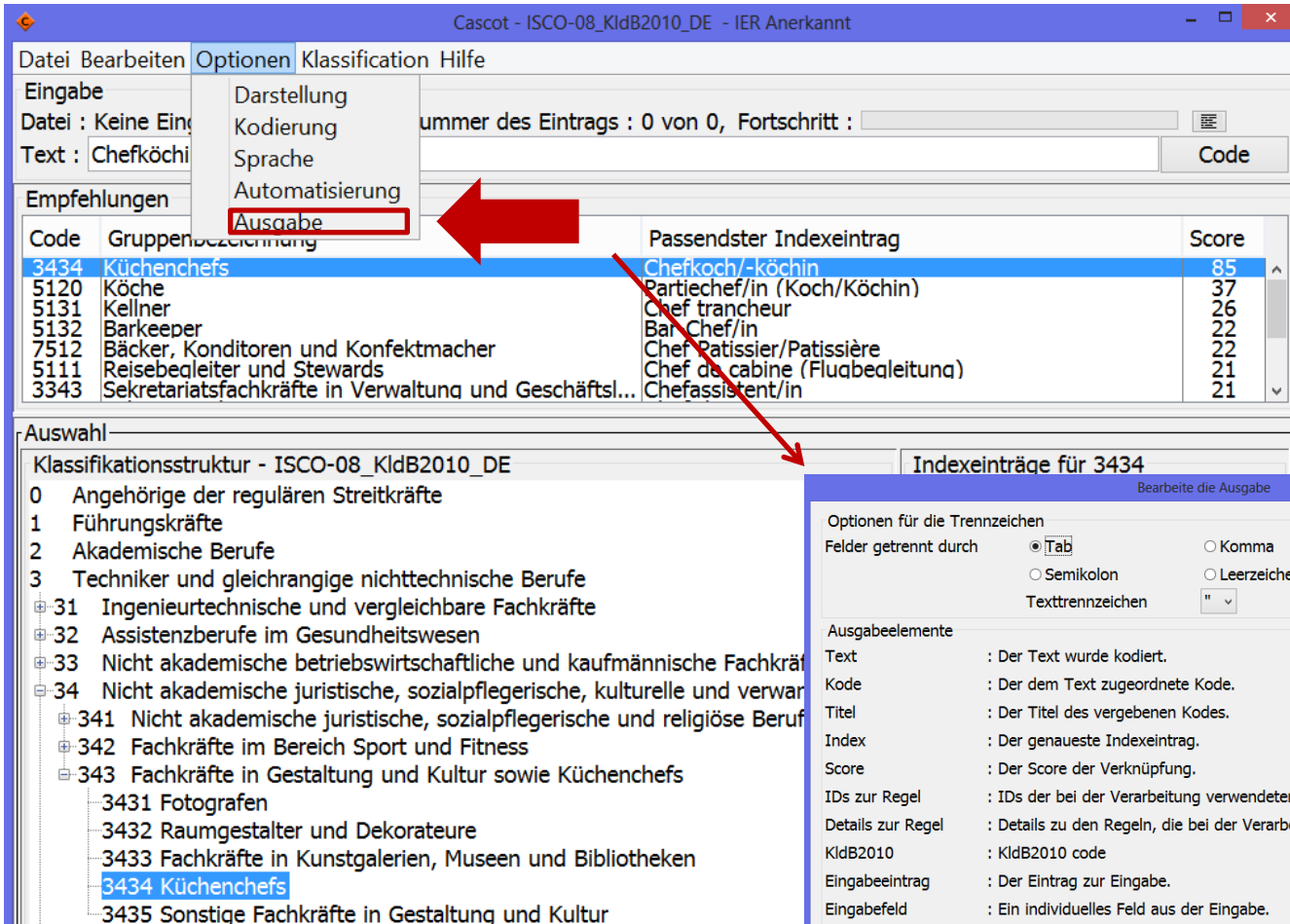
- 0 Armed forces occupations
- 1 Managers
- 2 Professionals
- 3 Technicians and associate professionals
- 4 Clerical support workers
- 5 Service and sales workers
- 6 Skilled agricultural, forestry and fishery workers
- 7 Craft and related trades workers
- 8 Plant and machine operators, and assemblers
- 9 Elementary occupations

**Indexeinträge Table:**

Indexeintrag	Score

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

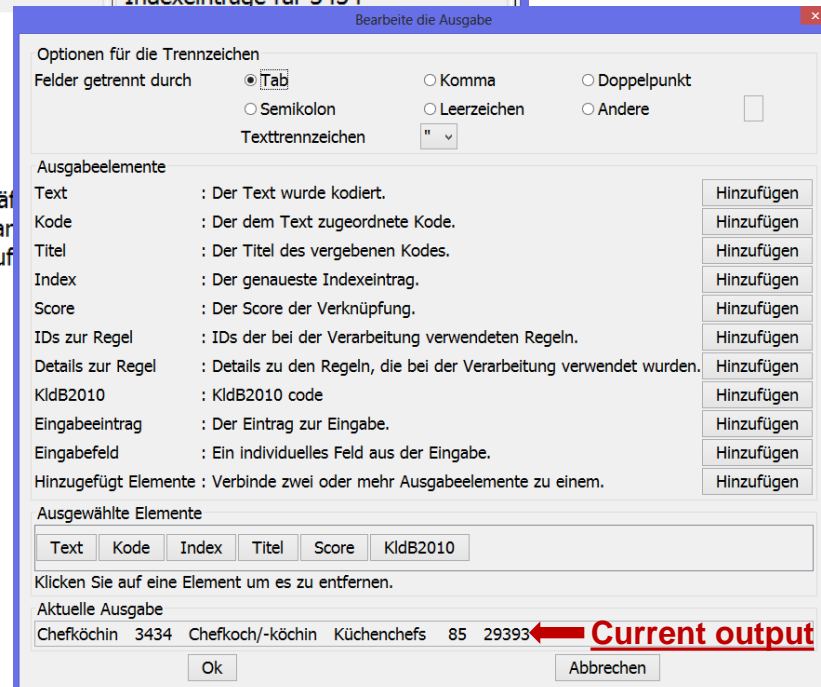


The screenshot shows the 'Cascot - ISCO-08\_KldB2010\_DE - IER Anerkannt' window. The 'Ausgabe' menu option is highlighted in the 'Optionen' menu. A red arrow points from this menu item to the 'Ausgabe' dialog box. Another red arrow points from the 'Ausgabe' dialog box to the 'Auswahl' pane, which shows a tree structure of classification items. The item '3434 Küchenchefs' is selected in the tree.

Code	Gruppenbezeichnung	Passendster Indexeintrag	Score
3434	Küchenchefs	Chefkoch/-köchin	85
5120	Köche	Partiechef/in (Koch/Köchin)	37
5131	Kellner	Chef trancheur	26
5132	Barkeeper	Bar Chef/in	22
7512	Bäcker, Konditoren und Konfektmacher	Chef Pâtissier/Pâtissière	22
5111	Reisebegleiter und Stewards	Chef de cabine (Flugbegleitung)	21
3343	Sekretariatsfachkräfte in Verwaltung und Geschäfts...	Chefassistent/in	21

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.



The 'Bearbeite die Ausgabe' dialog box is shown. It has two main sections: 'Optionen für die Trennzeichen' and 'Ausgabebelegte'. The 'Optionen für die Trennzeichen' section has radio buttons for 'Felder getrennt durch' (Tab, Komma, Doppelpunkt, Semikolon, Leerzeichen, Andere) and a 'Texttrennzeichen' dropdown. The 'Ausgabebelegte' section lists various output elements with 'Hinzufügen' buttons. A red arrow points to the 'Hinzufügen' button for 'Text'. Below the list, there are buttons for 'Ausgewählte Elemente' (Text, Kode, Index, Titel, Score, KldB2010) and a 'Klicken Sie auf eine Element um es zu entfernen.' section. At the bottom, there is a preview of the 'Aktuelle Ausgabe' and 'Ok'/'Abbrechen' buttons.

Aktuelle Ausgabe
Chefköchin 3434 Chefkoch/-köchin Küchenchefs 85 29393

**Current output**

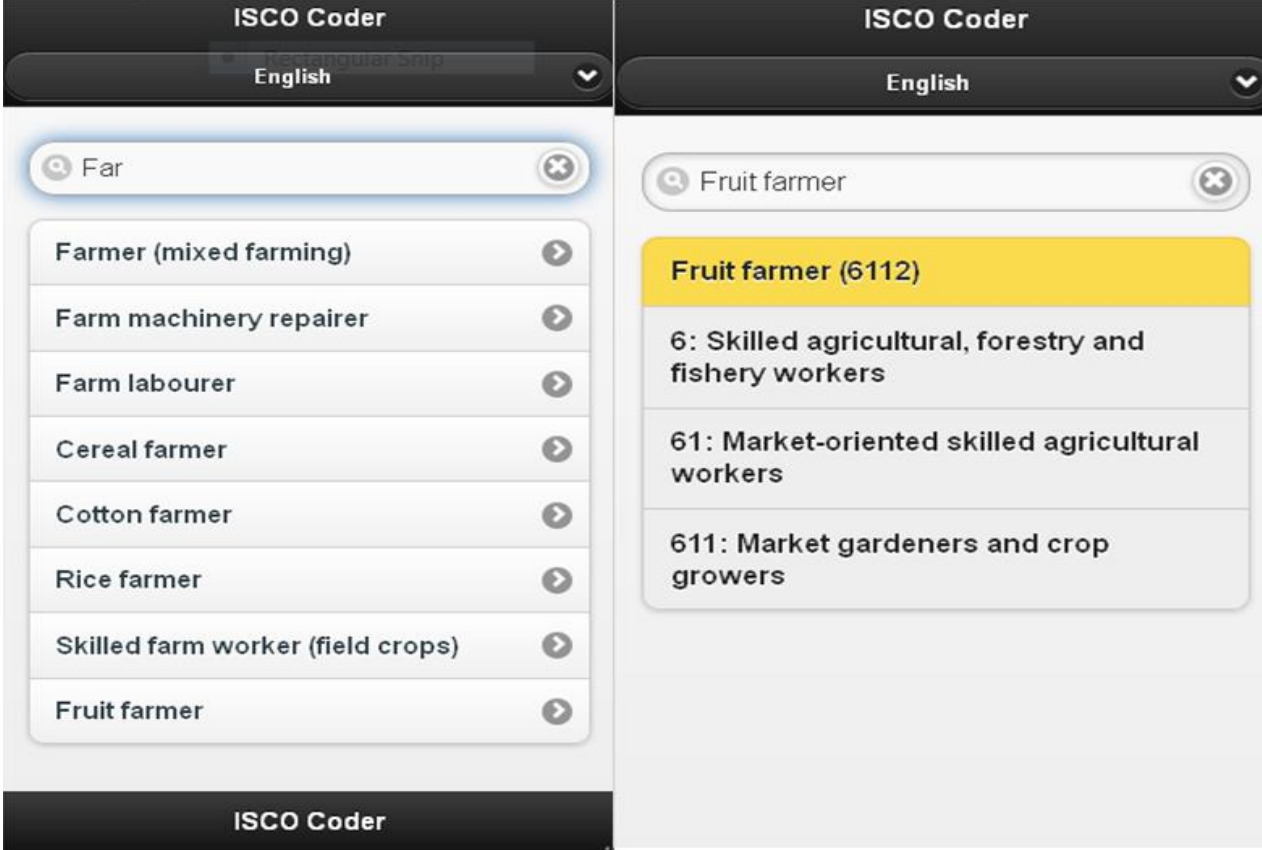
## Parallel work: CentERdata's Jobcoder app

- 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

## 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**

English

Far

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

**ISCO Coder**

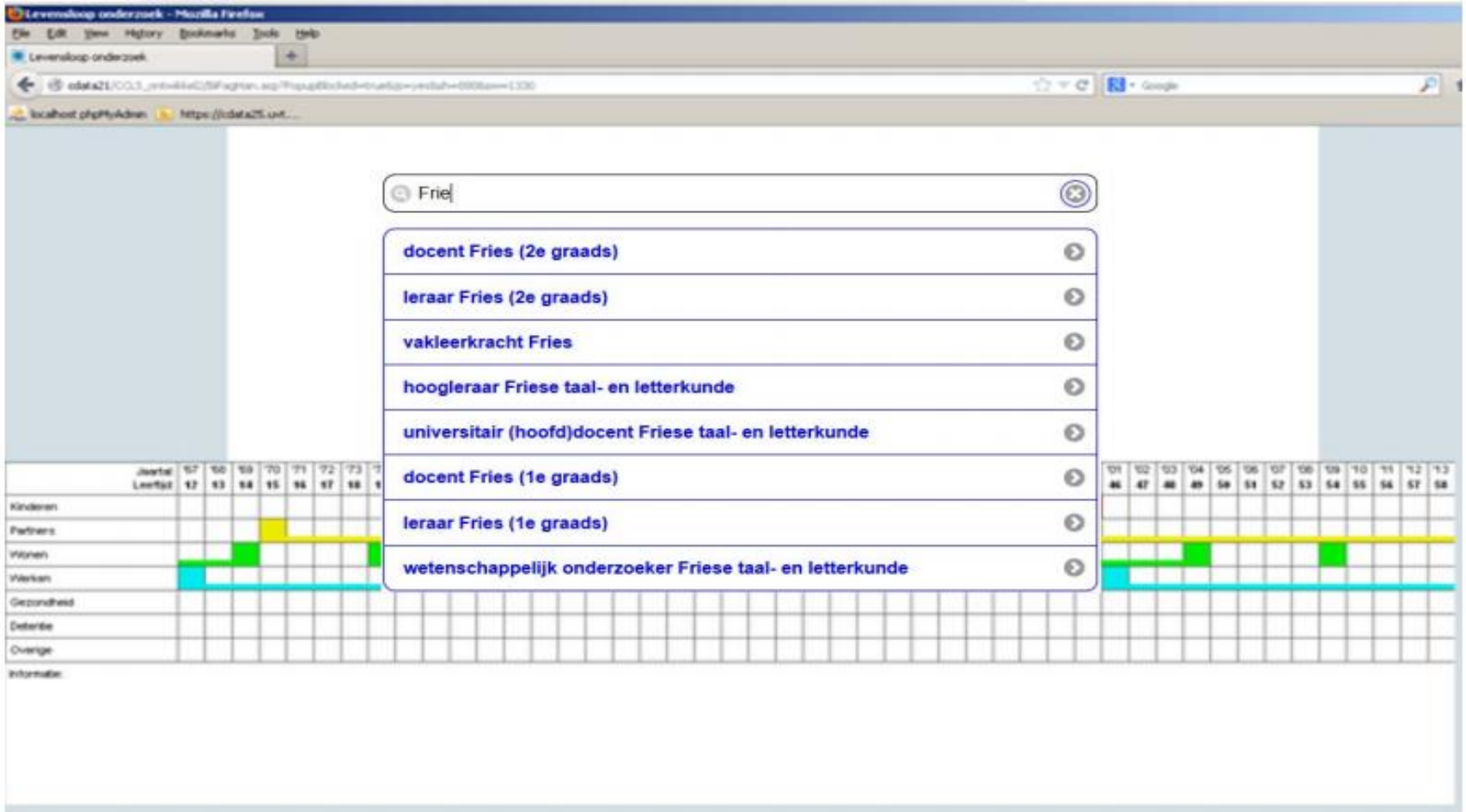
English

Fruit farmer

- Fruit farmer (6112)**
- 6: Skilled agricultural, forestry and fishery workers
- 61: Market-oriented skilled agricultural workers
- 611: Market gardeners and crop growers

**ISCO Coder**

# Jobcoder embedded in CAWI



The screenshot shows a web browser window titled 'Levensloop onderzoek - Mozilla Firefox'. The address bar contains a URL starting with 'odata21/003\_...'. The main content area features a search bar with the text 'Frie'. Below the search bar is a dropdown menu listing several job titles, each with a right-pointing arrow icon:

- docent Fries (2e graads)
- leraar Fries (2e graads)
- vakleerkracht Fries
- hoogleraar Friese taal- en letterkunde
- universitair (hoofd)docent Friese taal- en letterkunde
- docent Fries (1e graads)
- leraar Fries (1e graads)
- wetenschappelijk onderzoeker Friese taal- en letterkunde

Below the dropdown menu is a timeline chart. The x-axis represents years from 1997 to 2013. The y-axis lists various life events: Kinderen, Partners, Wonen, Werken, Gezondheid, Deterie, Overige, and Informatie. The chart shows colored bars indicating the duration of these events. For example, 'Werken' is shown in cyan from 1997 to 2008, and 'Wonen' is shown in green from 1997 to 2008. The 'Frie' job search results are overlaid on this chart, with the search bar and dropdown menu positioned over the 'Werken' and 'Wonen' bars.



## 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](mailto:Eric.harrison.2@city.ac.uk)

[www.europeansocialsurvey.org](http://www.europeansocialsurvey.org)