



Working Session 8: The UN Classification systems in the upcoming 2010 Census round



Measuring economic activities in the upcoming 2010 Census round

QUESTIONS, RESPONSE CATEGORIES AND LAYOUT





Question order and sequencing - general

- Questions on status in employment, occupation and industry are asked only of the economically active
 - Can be asked of both employed and unemployed (by reference to last job)
 - May be restricted to persons over a certain age
 - Questions are sequenced from question(s) on Economic activity status (Employed, Unemployed, Economically inactive)
- Questions on education are asked of all persons (over 5?) and usually precede those on economic characteristics





Economic activity status: hypothetical questions recommended for testing and use.

1 Did you/... do any work for one hour or more in the last 7 days?

2 Did you/... do any unpaid work in a family business or farm in the last 7 days?

3 Did you/... do any work at all in the last 7 days?

Prompts:

Any work on your farm or kraal?

Any fishing or seafood collecting?

Make anything for sale or your own use from farm or natural products?

Fetch any water or collect any firewood?

Any work in a business of any type?

Any type of wage job (full-time or part-time)?

4 Do you ... usually work but happened to be absent last week because of leave, sickness, bad weather, industrial troubles or other reasons?

Interviewer: If "yes" to any of 1 to 4, skip to next section.



Unemployment

5 Did you actively look for work sometime during the last four weeks?

Yes

No – seasonal worker awaiting busy season

No – believe no work available

No – tired of looking

No – don't know how or where to look

No – awaiting appropriate work

No – bad weather

No – family responsibilities

No – own illness

No – other reasons (specify)

6 If offered a work opportunity could you have started work last week?

Yes

No





Population not currently active

7 What were you doing/ what was your situation last week?

Studying	1
Performing housework.....	2
Not working and disabled.....	3
Sick.....	4
Retired/aged.....	5
Pension, rental or other income recipient.....	6
Other	7





Typical ordering of main questions on economic characteristics

1. Income (if asked)
2. Economic activity status
3. Hours worked/working time (if asked it usually relates to all jobs)
4. Status in employment/institutional sector in main job
5. Occupation
 - a) Title
 - b) Tasks
6. Industry
 - a) Kind of industry/activity
 - b) Name and address of employer



Were you self-employed or working for someone else in your (main) job last week?

- Self-employed
 - With paid help (employer)
 - Without paid help
- Owner/manager of incorporated business
 - With paid help (employer)
 - Without paid help
- Worked for someone else
 - As Government employee
 - As employee of a foreign Government
 - As employee of private company/person
 - As unpaid worker in family business/farm
 - Not stated





Occupation: Hypothetical questions recommended for testing and use (part 1)

**(In the main job held last week) what was (your) work or
occupation?**

Please give full job title and be specific, for example:

- Fruit picker
- Legal secretary
- Restaurant manager
- Secondary school teacher
- Cattle farmer
- Registered nurse

Occupation:

.....

.....





Occupation: Hypothetical questions recommended for testing and use (part 2)

What are your main tasks or duties in that job?

Please give details. For example:

Picking and carrying oranges and peaches

Preparing legal documents

Managing the operations of a restaurant

Teaching mathematics

Managing a cattle farm

Caring for the sick and administering medications

Main tasks or duties:

.....
.....
.....





Industry:
Hypothetical questions, recommended for
testing and use

What is the name and address of your employer or business?

(a) Name _____

(b) Address _____

**What are the main goods or services produced by your
employer or business?**



The UN Classification systems in the upcoming 2010 Census round

UN CLASSIFICATIONS AND CODING TOOLS AND THE USE OF TECHNOLOGY IN THE CENSUS





Coding occupation and industry

- Responses to open ended questions have to be assigned to the appropriate category in an occupation classification and in an industry classification
- Not a simple process
- Responses to questions on occupation (title and tasks), industry and name and address of workplace are relevant to both coding processes
- Coding should be done using indexes of occupations and industry
 - Mapping directly to the classification is error prone and inefficient





Information needed before planning occupation and industry coding processes

- What classifications of occupations and industry are used in other surveys (LFS, employer surveys) and in national employment services?
- Who is responsible for these classifications?
- Are these classifications based on ISCO-08 and ISIC Rev 4?
- Are there plans to update the national classifications?
- Is there a national index of occupations?
- Is there a national index of industries?
- What user needs impact on decisions about coding:
 - Is there a need for information about small occupational groups for specific small geographic areas?
- What happened in the last Census?





Occupation and industry coding as part of the Census processing strategy

- The main aim of the coding process
 - To determine and record correctly to which of the categories in the respective classifications the jobs belong
 - at the most detailed level of the classification possible on the basis of the information provided in the responses
 - task to be completed within an overall processing plan for the census
 - to a pre-specified timetable
 - within pre-specified cost limits or in a fashion that will minimize cost, given the specified data requirements.
- Important factors to consider
 - The existing data processing capacity and infrastructure
 - The type and format of the information to be processed
 - The volume of data to be processed and the throughput rates required
 - How processing of industry and occupation is embedded within the total data processing task for the census
 - The level of detail required to satisfy important user needs in the national context, as well as for international reporting





Strategic coding and processing options (1)

- Process all cases or a sample only
 - Implementation at collection stage (long and short forms) or at processing stage
 - Process occupation for only a sample of the data collected
 - Ethical concern about collecting data that are not used
 - Process a sample for early release
 - Significant cost and time savings
 - Data for small areas, small populations and small occupation and industry groups may be severely compromised





Strategic coding and processing options (2)

- ***Field or office coding?***

- The following choices are available:

1. The respondent codes himself/herself to a predefined category

- Implies pre-coded questions
- Inadequate quality but less expensive than other options

2. The enumerator codes in the field, either during the interview or before the questionnaire is forwarded for further processing

- A variation on pre-coded questions

OR

- Enumerator writes down the response (or keywords) and codes the response after the interview using an index.

3. Specially trained coders code in connection with consistency checks of the questionnaire and data entry

- Office coding – the preferred option in most cases





Coding in the field by the enumerator

- Enumerator writes down the response (or keywords) and codes the response after the interview using an index.
 - Allows coding to detailed categories
 - Over time and as part of training enumerators become aware of the type of information required to code correctly

BUT

- Coding becomes one of many tasks for a large number of enumerators
- Enumerators cannot be given the same amount of training, supervision and support as specialized coders
- Quality and consistency of coding are likely to be poor
- This approach can work in a continuous Labour Force or similar survey with a permanent field staff and good training and communication
- For large-scale infrequent operations such as the Census, coding of occupation and industry as part of the central processing operation is preferred





Office coding

- Can be done manually with a paper index, or with computer assistance
- Can be combined with automatic coding
- Coders may specialize in the coding of one (or a few) variables or deal with the whole form
- Coders need to be thoroughly trained and tested before they start coding
- Quality of coding operations can and should be rigorously controlled





Office coding – common problems and strategies to avoid them

- Common problems

- Coding teams may become production minded and focus on throughput at the expense of quality
- Coders (individuals and groups) may find short cut methods that introduce systematic errors
- ‘Memory’ coding
 - Training, supervision, quality control (eg. recoding a sample), audits, workshops on coding problems, index updates for common un-codable responses
- Boredom with a repetitive process
 - Rotation between processes, good job design
- Poor motivation of staff towards end of contract
 - Motivation schemes, offer long term employment to strong performers
- Coding centre and organization may be overwhelmed by volume of documents at beginning of process
 - Good planning, dress rehearsal should include coding





At what level of the classification should responses be coded?

- In past censuses it was frequently decided that responses should be coded to an aggregate level of the classification structure (Eg 3-digit level of ISCO)
- Arguments for this include:
 - Perceived cost of coding to a larger number of categories in terms of errors and staff hours required
 - It was thought that the responses would not support coding to more detailed categories
 - When coding only a sample the detailed results may not be publishable



At what level of the classification should responses be coded?

- But the experience of statistical agencies has shown
 - Marginal costs of coding to a larger number of categories are small
 - Error rate does not significantly increase and may improve for aggregate groups
 - Many responses support detailed coding, while some do not
 - Coding to an arbitrary specific level of the classification involves unnecessary loss of information
 - Some groups at detailed levels of classification may be larger than others at higher levels
 - Coding at a higher level limits options for tabulation, international reporting, production of flexible non-standard aggregations





Coding of vague and difficult responses

Some responses may be too vague and imprecise to allow the coder to determine to which category the job belongs

- These responses should be coded to the level in the classification structure supported by the information contained in them
- should not be forced into any particular detailed category where only a small proportion of the jobs would fall if the responses were adequate.
 - For example, in one census 15 per cent of the jobs coded to the major group “clerks” could not be coded to any of the more detailed categories within that major group.
- Residual groups (not elsewhere classified) should not be used for vague responses
- A common method of dealing with this type of response is to provide entries in the coding index for commonly occurring vague responses
- Such responses are assigned the code for the relevant higher category, followed by trailing zeros.
 - responses can be allocated proportionally to the more detailed categories in a transparent manner
 - or they can be released in publications labelled as : ‘Group name not further defined’





Occupation title: Sales

Tasks performed: Selling

52 Sales workers

5200 Sales workers not further defined

521 Street and market salespersons

- 5211 Stall and market salespersons
- 5212 Street food salespersons

522 Shop salespersons

- 5221 Shop keepers
- 5222 Shop supervisors
- 5223 Shop sales assistants

523 Cashiers and ticket clerks

- 5230 Cashiers and ticket clerks

524 Other sales workers

- 5241 Fashion and other models
- 5242 Sales demonstrators
- 5243 Door to door salespersons
- 5244 Contact centre salespersons
- 5245 Service station attendants
- 5246 Food service counter attendants
- 5249 Sales workers not elsewhere classified





Other difficult responses

- Indicate that the establishment (or job) in question produces a combination of goods or services (or the job involves tasks and duties) that cut across the distinctions made in the industry or occupation classification
- Represent a type of production or work not covered by the classification or not listed in the coding index
- Should be referred as queries to expert coders and ultimately to those responsible for maintaining the classification and index
- Strategies to deal quickly with commonly occurring difficult responses and update the coding index or procedures are essential





Key decisions to make and questions to answer in planning the coding operation

- Which classifications will be used?
- Do you need to update the classification or index before the Census process?
- Level of coding?
- Field or centralized coding?
- Specialized teams for each topic?
- Manual, computer assisted or automatic coding
- What levels of error can you tolerate?
- Coding throughput rates?
- How many coders do you need and for how long?
- **How much will it cost?**





Types of industry coding index

- Most census coding operations will find it useful to have two coding indices for the coding of industry, as follows:
- A list of as many as possible of the establishments operational in the geographical region covered by the coding operation
 - each establishment has been given the correct industry code by specialists in establishment surveys and in the coding of establishment activity.
 - usually cover only large, formal sector establishments as they have been created from lists kept in tax offices, licensing offices and/or chambers of industry and commerce
 - may nevertheless cover significant proportions of the work force, and their use for census coding will eliminate one possible source of inconsistency in employment statistics between the census results and the results of establishment surveys;
- A list of significant word combinations reflecting the answers given in response to industry questions
 - an index of the same type as that created for the coding of occupation.





Industry coding index and process

- The industry coding process will usually attempt first to match the name and address of the person's employer with those in the list or register of establishments.
- If a match cannot be made using the register of establishments, then an attempt is made to match the description of the industry with the index of type (b) above.
- The process of updating the coding indices for industry responses should be viewed as part of the general process required to maintain the industry classification
- Both types of index will need to be updated in advance of the Census and modified for coding of responses given in household based surveys
- Experience shows that only a minority of census responses will be coded successfully using a list of establishments





Sources of information for developing an industry index

- Reviews of responses from recent survey operations and census tests
- Business registers and directories
- Notices and advertisements in newspapers, journals, and on the internet
- ISIC index may be a good starting point in the absence of any national index or a useful source to assess completeness of a national index
- Do neighbouring countries with similar languages have indexes you could share or adapt?
- Index needs to reflect language as used in the country





Organization and structure of the industry index

- As with occupation the index may be all inclusive or structured

Response:	Keyword/first qualifying word/second qualifying word:
Sheep farm:	sheep/farm
Car rental agency:	car/rental
Youth club:	club/youth
Tax assessment office:	tax/assessment office
Cleaning service:	cleaning/services
Cleaning products production:	cleaning/products/production



Structure of the industry coding index

- The keyword is the word in the relevant response that alone can serve as a designation of a service, a product or a function, however imprecise.
- The qualifying words will usually indicate some special form or variety and/or the type of activity associated with the product or service.
- That sequence has been chosen because the number of different designations for activities is much smaller than the number of designations for different products, services and functions.
- Sometimes the keyword may be precise and in itself suffice as an index entry, such as “abattoir”.
- However, the keyword may also be very ambiguous, such as ‘advertising’.





Automatic and computer assisted coding

- Both approaches use computing power to speed up process of searching an index, identifying matching responses, following coding rules and recording the correct code
- In Computer assisted coding (CAC) the coder enters a small number of characters from key and qualifying words
 - Matching index entries are displayed and coder selects matching entry
 - Correct code is recorded by coder or the computer or a query is raised
- In automatic coding (AC) responses are key entered or captured photographically, then matched automatically by the computer
 - Match rates of up to 70% have been achieved
 - Remaining entries are coded using CAC
 - Requires a high degree of sophistication and a very well designed index
- Software solutions are available at low cost but cost of integration into larger processing systems may be high

