Processing of Census of Population and Housing 2014: Scanning and Editing Case of Morocco

M. Ouassama MARSELI,
Chief of implementation and processing of Census of Population and Housing
omarseli@gmail.com
Table of content

- I. Introduction
- II. Publication strategy
- III. Processing steps
I. Introduction: quantitative indicators

- 9,566,000 forms (A3+A4)

- 75 questions (28 individual questions, 47 households questions).

- 7 questions with national classification:
  - Profession (1): 11,315,401 labels with 892 professions
  - Industry(1): 10,301,176 answers with 650 industries
  - Diploma (2): 13,867,209 answers and 1,638 diplomas
  - Migration (3): 26,657,176 answers and 1,838 communes
    - Which is about 62,140,962 labels
I. Introduction: qualitative indicators

- 81% of households with at least one error
- 41% of persons with at least one error
- 5.8 errors on average per household
- 1.1 errors on average per person
II. Strategy

- Objective: publish results on time as planned
- Phase 1: publish population count
- Phase 2: Publish variables pre-coded on the field
- Phase 3: Publish results of variables with classifications
III. Processing steps

- Step 1: Scanning and Intelligent Character Reading (ICR)
- Step 2: Interactive editing
- Step 3: Automatic editing
- Step 4: Manual codification assisted with computer
Step 1: Scanning and intelligent character reading

- Scanning with high speed scanners
- Intelligent character reading
- Quality control (ISO)
Step 2: interactive editing
Data editing objective

I. Identify and quantify errors in order to improve next census by reducing errors at the source (i.e. data collection phase)

II. Provide statistics on data quality to users

III. Identify and edits missing and erroneous data

IV. Provide complete and coherent data
Step 2: interactive manual editing

Control rules

- Control of completeness at the census district level, housing units, households, and persons (Structure control of omissions and duplications)

- Validity control (105)
  - Ex: matrimonial status non valid

- Person consistency control (87)
  - Ex: Fertility recorded for male (consistency)

- Household consistency control (14)
  - Ex: several household heads

- Housing units consistency control (46)
  - Other sources of electricity not recoded for housing units with no public power network
Step 2: interactive manual editing

- Editing instruction
- Application (in house)
  - Access to images of questionnaires
  - Verification of control rules
- Two levels of editing
  - Operator
  - Supervisor
- Errors decreased by 62%
**Step 3: Automatic editing**

**Principles:**

1. Satisfy all *controls* by changing the fewest possible items

2. Imputation algorithm should not use explicit *editing* rules

3. Verify that the *frequency* distributions of variables is maintained post-imputation
Step 3: Automatic editing

Method:

- Imputation of non valid and inconsistency items, persons and households

- Dynamic imputation using nearest neighborhood, which satisfies all control rules

  - Ex: Imputation of age using sex and relation to head of household
Step 4: Coding

- Questions requiring classifications are not coded on the field
  - Diploma
  - Profession
  - Industry
  - Migration

- Interactive coding: Key from image using look up tables of National classifications
Thanks