Experiences with Blaise for data editing in the housing rental survey

Harm Melief
Mark Kroon
Overview

1. The Housing Rental Survey
2. The system
3. Lessons learned
4. Conclusions
The housing rental survey

- (Partly) commissioned by the Dutch Government
- Fixed panel survey of approximately 14000 rented houses from 1300 respondents
- 2/3 paper forms, 1/3 register
- 50 variables of which 30 exclusively for the Government
- Data editing: within 4 weeks
  - Correcting omissions and mistakes
  - Performed by calling respondents
- Output:
  - Micro-data for internal use and for the government
  - Price index of rental costs → 15% of CPI
Support system specifications

- Data I/O: importing the forms, exporting cleaned data
- Data editing
  - Work distribution
  - Editing single houses
  - Editing groups of houses
- Price index calculation

- Two operational demands
  - Data storage in SQL
  - End-users do most of the maintenance → Keep it simple
System: Overall design

- System based on three database tables
- From these tables a number of views are made combining the three tables as needed

Diagram:
- Respon-dents
- Panel
- Editing

RespID

HouseID
Overall design

- Raw data
- Respondents
- Panel
- Data editing
  - Work distribution
- Editing
  - Index calculation
  - Export microdata
- Output

Panel management
Work distribution

Main controls: buttons

Forward

Work: Lookup

Workers

Filter: controls/buttons

Centraal Bureau voor de Statistiek
Data editing (1)

Panel

Respondent

Editing

Private information

Group of houses

History of selected house

DEP
Index calculation

- Rental prices:
  - 4 categories giving 64 strata (cells)
  - Within a stratum unweighted arithmetic mean
  - Final index calculated as weighted mean over all strata
- Service costs:
  - No strata used
  - Index calculated as unweighted mean over all houses
- Performed in Manipula
Lessons learned: Blaise

• Overall we liked Blaise

• Problems:
  • Categories are hard coded → maintenance
  • Performance:
    • Externals
      ➢ Solution: pre-selection external data
Lessons learned: Manipula

- Good ETL tool
  - Fast, especially with text files
  - Many different file types supported
- Problems:
  - Both input and output datamodel needed → many datamodels → maintenance
  - Performance database:
    - Joining slow, solution: views in database
    - Version control slow, solution: turned off
    - Importing data slow, solution: batch
Lessons learned: Maniplus

• Decent GUI builder
  • Controls support Blaise datamodels
  • Lookups powerful tool for presenting data

• Problems:
  • Lookups were buggy: new release
  • Maintenance:
    – Most GUI objects and their properties hardcoded
    – No wizard
Conclusion

- We managed to build a complete work scheduler and data editing tool in the Blaise Suite
- SQL for data storage
- Some concessions were needed because of limitations of the tool
- Performance was an issue

Blaise allows rapid building of a data editing tool. When higher complexity/performance is needed, its limitations should be taken into account.