Business Case for Industrialisation in Statistics Estonia: Small Example of a Large Trend

MSIS 2013
Allan Randlepp
Tuulikki Sillajõe

Outline

- Business case for Population and Housing Census 2011 (PHC 2011)
  A. Data collection phase
  B. Data processing phase
- The production system as a whole
- Conclusions and further developments

19.10.2010

statistics estonia

PHC 2011 Data collection phase: Mixed mode was used for PHC 2011

- Modes of questioning
  - e-Census on the web (CAWI) → 66%
  - Interview (CAPI)
- Institutions filled in a special questionnaire
- Different data sources
  - Population and Housing Census 2000
  - Administrative registers
  - Data collection from persons
- New software was developed next to eSTAT

19.10.2010

The costs of data collection of PHC 2000 and PHC 2011

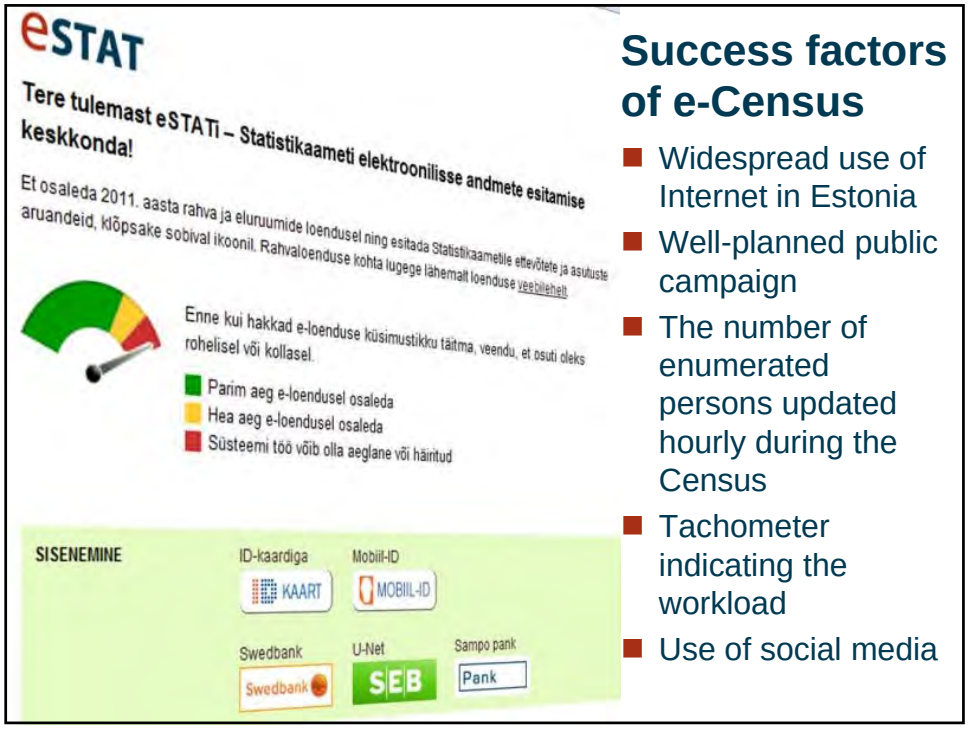
	e-Census participation rate	Number of enumerators	Cost of data collection, euros	Cost of data collection per enumerated person, euros
PAPI (prediction)	0%	4,592	17,710,000	13.2
CAWI+CAPI (prediction)	25%	2,970	10,910,000	8.14
CAWI+CAPI	66%	2,131	8,410,000	6.50

19.10.2010

The main reasons for lower costs of data collection during PHC 2011, compared to PHC 2000:

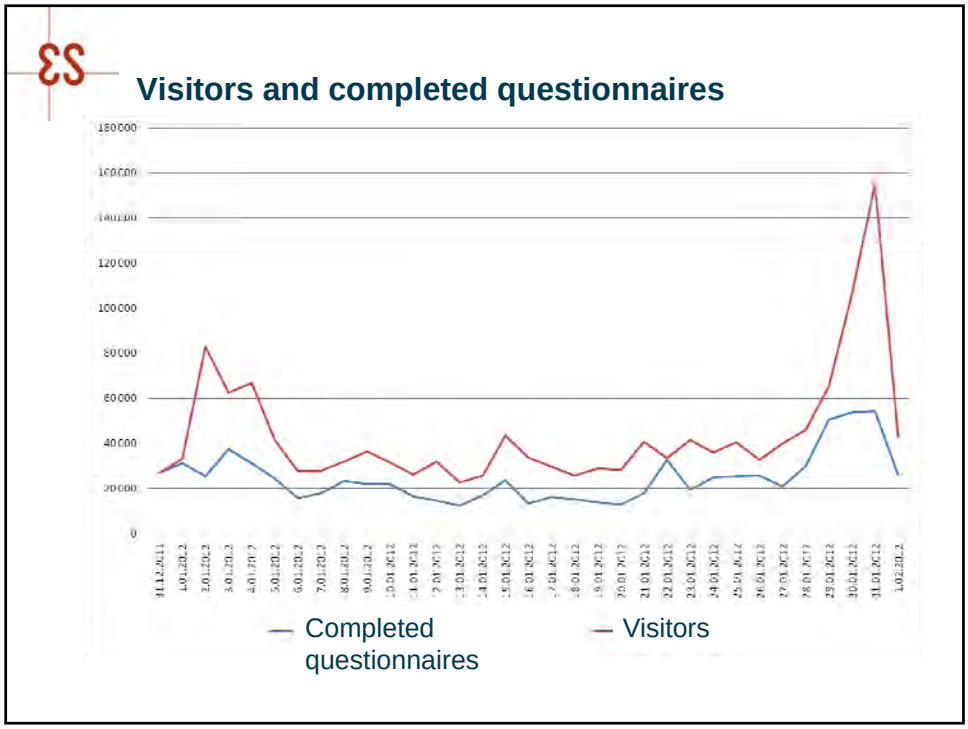
- There were no special data entry costs;
- Automatic checks diminished mistakes during interviewing in both cases (CAWI and CAPI);
- Less time was required for interviewing than planned, and therefore fewer enumerators were hired;
- Printing, communications and archiving costs were considerably smaller;
- Management of data collection was much cheaper because much fewer employees were needed, thanks to the management module of the new software.

19.10.2010

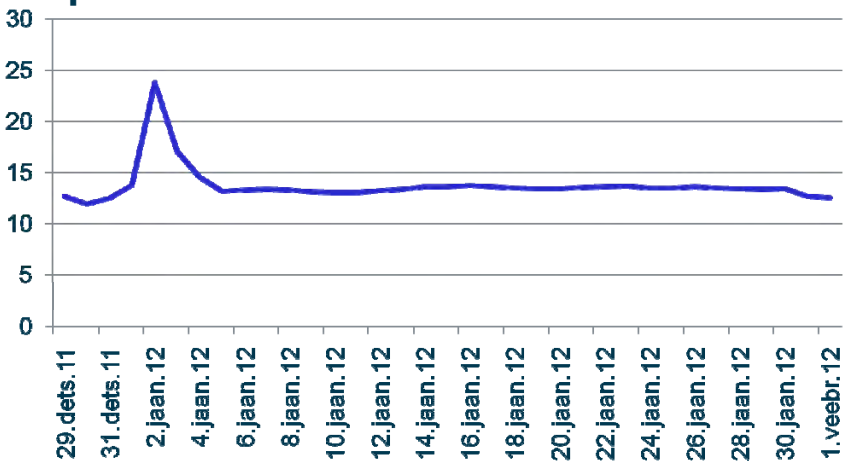


Success factors of e-Census

- Widespread use of Internet in Estonia
- Well-planned public campaign
- The number of enumerated persons updated hourly during the Census
- Tachometer indicating the workload
- Use of social media



Average completion time of questionnaire



Data processing phase

Costs of data processing of the PHC 2011

	e-Census participation rate	Number of operators	Cost of data collection, euros
PAPI (prediction)	0%	264	3,600,000
CAWI+CAPI (prediction)	25%	90	2,300,000
CAWI+CAPI	66%	30	1,090,198

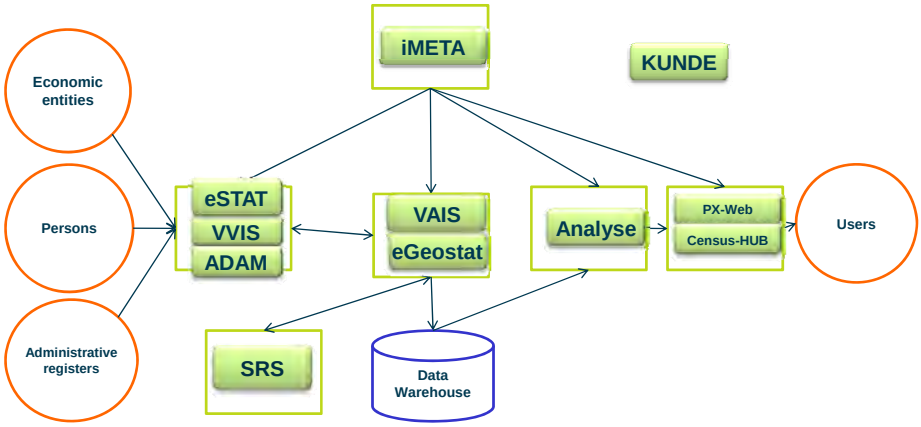
19.10.2010

The main reasons for lower costs of data collection during PHC 2011

- Automatic checks diminished mistakes during interviewing in both cases (CAWI and CAPI);
- The first stage of data processing, primary data arrangement, was part of the data collection software and ran as parallel activity during data collection and immediately after collection. As a result the collected data was prepared for the next stages of data processing.
- Data processing was highly automated. Totally 35.7 million errors were corrected, 20.2 million of them corrected automatically and 15.5 million manually.
- Problems with data came out in early phases of processing, because statisticians were able to monitor data during the data processing.
- For checks and handling missing data was possible to use data from registries and from PHC 2000.
- Management of data collection was much cheaper because much fewer employees were needed, thanks to the automated data processing done by the new software.

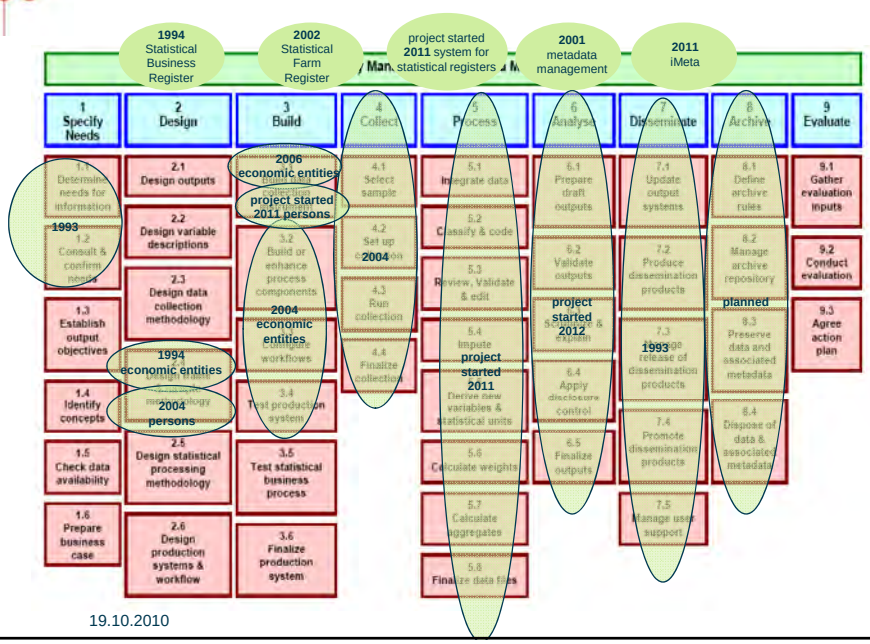
19.10.2010

Architecture of the information system



19.10.2010

Generic Statistical Business Process Model



19.10.2010

Sources for efficiency gains of Statistics Estonia

- Central data collection department, i.e. standardisation of processes
- Generic office-wide software
- Administrative data instead of survey data if applicable
- Pre-filling of statistical questionnaires with administrative data

19.10.2010

Small developments, big efforts

- Reminders sent before the deadline instead of after the deadline
- Informing economic entities simultaneously about all the questionnaires they have to fill in next year
- Centralisation of the preparation of questionnaires from statistical departments to IT Department and within a few years to Methodology Department
- Standardisation and simplification of instructions about questionnaires for economic entities
- Creation of a list of input variables

19.10.2010

Next practical steps

- Introduction of CAWI as the main data collection method for surveys on individuals (2013)
- Introduction of CATI for data collection from both types of respondents: economic entities and individuals (step by step, starting from 2012)
- Implementation of generic office-wide software for other functions than data collection
- Reuse of data within the statistical office

19.10.2010

Strategic directions

- Training of data suppliers (economic entities, individuals, registers, etc.), incl. about their personal and public gains
- Closer cooperation between the data collection function and dissemination function within the organisation, for better communication with data suppliers (based on the experience of PHC 2011)
- Simplification of statistical reports (harmonisation of concepts, deadlines, practices, etc.)
- Development of infrastructure for selling data collection services

19.10.2010

Further challenges

- Wider use of administrative and commercial data
- Do we need two data collection tools?
- Centralization of data processing function?

19.10.2010