**UNITED
NATIONS**

E

Economic and Social Council

ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Group of Experts on Population and Housing Censuses

Twelfth Meeting
Geneva, 28-30 October 2009
Item 6 of the provisional agenda

ON LINE DATA COLLECTION

An online system for multi-channel, register-based census data collection

Note by the National Institute of Statistics, Italy

I.   THE DATA COLLECTION STRATEGY FOR THE FIFTEENTH GENERAL POPULATION CENSUS

1.   The collection strategy for the 15th general Population Census is based on the use of municipal registers for the identification of survey units. The shift from a *traditional census* to a *register-assisted census* will be supported by new collection techniques; questionnaire mail-out and a variety of possible data return methods (post, web compilation, and delivery to the municipal collection centres). Municipal address archives, geocoded to the census tract, will also be used.

2.   Another fundamental feature of the new strategy is its *modularity*, that is the adoption of different methods and techniques in relation to the demographic size of the municipalities. The classes are listed in the following table.

Table 1
Number of municipalities and 2001 population by demographic size

Type	Demographic size	Number	2001 population (%)
A	Province capitals and municipalities with at least **50,000** inhabitants	165	35.8
B	Municipalities with between **20,000** and **49,999** inhabitants (excluding province capitals)	339	17.0
C1	Municipalities with between **5,000** and **19,999** inhabitants (excluding province capitals)	1.859	29.6
C2	Municipalities with fewer than **5,000** inhabitants	5.738	17.6

3. The following modules will be part of the new census strategy for all municipalities:

    (a) **use of municipal registers to guide the collection** together with the use of new questionnaire distribution and multi-channel collection techniques;

    (b) **retrieval, by enumerators, of missing responses**, that is questionnaires not returned by respondents listed in the register;

    (c) **retrieval, by enumerators, of register under-coverage**, that is the identification of survey units not listed in the registers (people normally living in the municipality but not yet registered);

    (d) **survey-contextual census-register comparison.**

4. Other modules (census tract definition, civic number enumeration) are intended to support further innovations concerning only type A and B municipalities. In the latter, municipal address archives, geocoded to the census tract and prepared by the Italian National Institute of Statistics (Istat), will be validated by **civic number enumeration (CNE)**, with the aim of integrating national archives with any local municipal archives. In addition, a **dwelling census** will be performed to list the number of potential housing units for each civic number. This information, together with the civic number archive, will be used by enumerators to guide the missing response and under-coverage retrieval activities.

5. The *short form / long form* **strategy** will also concern only type A and B municipalities. It will be used to estimate representative socioeconomic variables at the *census area* (census tract clusters) level.

6. In the pilot census to be performed at the end of October 2009, the questionnaire length will be tested by proposing three form types: short, long and medium. The latter is an alternative to the short form.

7. In summary, **in type A and B municipalities** the new strategy will consist of the following modules:

    (a) census tract update;
    (b) census area creation;
    (c) development of a prototype civic number archive;
    (d) civic number enumeration;

(e)   register-assisted census;
(f)   short form / long form strategy;
(g)   missing response retrieval;
(h)   under-coverage retrieval;
(i)   **census**-register comparison.

8.   **In the type C1 and C2 municipalities** the new strategy will consist of the following modules:

(a)   census tract update;
(b)   register-assisted census;
(c)   *short form / long form* strategy (for C1 municipalities only);
(d)   complete census tract reconnaissance by the enumerator, based on information from the collection management system and aimed at retrieving missing responses and under-coverage and carrying out the dwelling census;
(e)   census-register comparison.

9.   The enumerators' work should be made easier and lighter than in past censuses by the use of **information** from the **management system**. This information will, for example, enable their daily route to be restricted to the addresses of those households that have not yet returned the questionnaire.

10.   The pilot census will be performed on a stratified sample of about 80,000 households and will allow all the described methods to be tested, with the exception of the civic number enumeration and the census-register comparison. It will also enable us to test the municipalities' use of the management system.

II.   THE ARCHITECTURE OF THE WEB-BASED SYSTEM

11.   The system for the pilot census for the 15th population census comprises two sub-systems: the data collection system and the census management system.

12.   The data collection system is a web application providing: credential-based access to households; online questionnaire completion; rule-based automatic data validation; upload of completed questionnaire; display and saving on the respondent's PC of a printable image of the completed questionnaire, as proof that it has been uploaded; automatic replication of final data in the management system.

13.   The management system allows the census network organisations to monitor and manage all field operations. The described innovations need some additional functions over an ordinary monitoring system: synchronisation and data exchange with external systems; synchronisation with the web-based collection system; higher complexity of monitoring functions due to diversification of questionnaires and their multi-channel return to municipalities; writing rights to be enabled for a higher number of user categories; explicit management of information on the status of each questionnaire; use of information to guide enumerators' daily and systematic activity.

14.   The monitoring system, in its roles as a distributed information and communication system for the data collectors and operational management system, thus becomes the fundamental element of the census machine. It must therefore provide the collection network with all necessary services: accessible, useable management and update functions and summary reports updated in "real time" by input information.

15.   The system for the pilot census will provide the 31 sample municipalities, Istat operators and other census network operators with the above listed functions.

16.   Each sub-system uses its own relational database and communication between them at the data level is one way; that is, the collection sub-system feeds the management sub-system by copying the final data once they become available. This choice, made for the pilot census prototype, has the following advantages:

(a)   sub-system independence;
(b)   separated performance control and differentiated tuning;
(c)   improved data privacy protection.

17.   It is worth noting that this approach will have to be re-discussed if the Census online questionnaire infrastructure is outsourced.

18.   The following figure depicts the general system architecture and sub-system interconnection.

Figure 1. The web system architecture
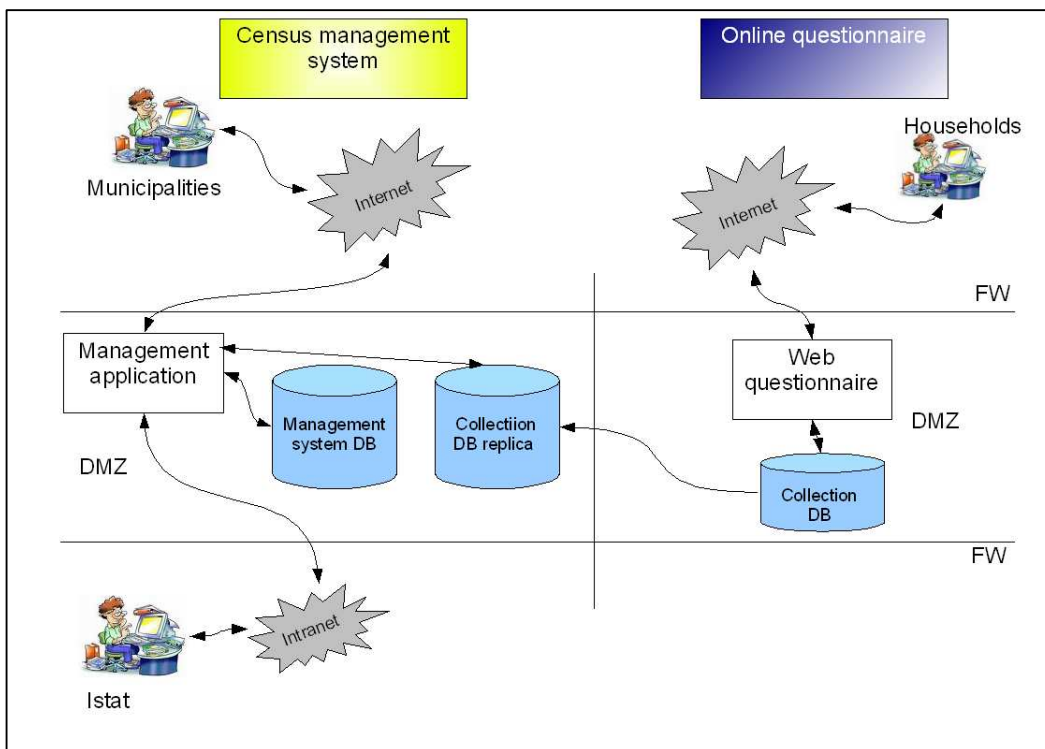
III.   THE SURVEY MANAGEMENT SYSTEM

19.   The Survey Management System allows management of enumerator activity and the production and provision of questionnaire delivery and collection information.

20.   System access will be controlled by a module allowing user profiles and consequent diversification of authorisations to be defined according to the responsibilities of each user, including at different territorial levels.

21.   The profiles in the prototype are the following:

(a)   ISTAT – Central and Regional officers;
(b)   Census network organisations – Back Office of the Municipal Census Bureau (MCB);
(c)   Other national organisations.

22.   The following functions have been implemented in the prototype sub-system:

(a)   census tract status control;
(b)   survey progress monitoring;
(c)   operator management;
(d)   various questionnaire-related operations;
(e)   document provision.

A.   Census tract status

23.   The census tract status control is the central function of the management system. It allows guidance of field operations through the census tract agenda and supervision of the collection result for each household from the municipal register through summary and detailed auxiliary forms. This function can be used by any enabled operator type (enumerator, coordinator or municipal manager).

24.   The census tract agenda keeps track of changes in the status of each questionnaire. It is initially populated with data from the population register and the list of all addresses of each census tract. The agenda is continuously updated with data from the collection system and input from the enumerators and back office operators.

25.   Summary and detailed auxiliary forms provide information on housing type and household composition. The detailed form contains information from a single questionnaire. It is completed by the collection system directly, if the questionnaire came through the Web, or by a municipal operator if a paper questionnaire has been received. The summary form provides territorially aggregated information, from the census tract level up.

B.   Summary reports

26.   Summary reports comprise the management system's collection monitoring section. They allow an "almost-real time" evaluation of the progress of field operations.

Collection control must be performed at the various territorial levels, based on users' responsibility and visibility permissions. In consequence, summary reports can be browsed with the *drill-down* technique from the regional to the census tract level.

27.   The reporting functions produce aggregate data tables on the number of returned questionnaires broken down by:

    (a)   type of questionnaire (short, medium or long form);
    (b)   type of returning channel (Web, post, municipal centre, enumerator).

28.   Reports on the collection status from the operations starting time are also available. A municipal manager will be able to view reports for all pilot census tracts, while enumerators may only view those census tracts they have been assigned and coordinators only those tracts assigned to the enumerators they supervise.

C.   Operator management

29.   This group of functions allows municipalities to manage back and front end personnel autonomously. Users enabled to these functions can enter, visualise and modify coordinator and enumerator personal information. Once a new operator has been inserted, the system creates his/her personal credentials, which can be communicated to the operator by the system directly via SMS or email.

30.   Enabled operators can also assign enumerators to coordinators and census tracts (and questionnaires) to enumerators and reset enumerator and coordinator passwords as needed.

D.   Questionnaires

31.   This group of functions allows the personal information printed on the questionnaires sent to respondents and on those returned via the Web to be viewed. It also possible to print empty questionnaires and additional personal forms. In this section, a municipal operator can compare the register data with census results. The report provides information on each individual questionnaire.

IV.   THE WEB-BASED QUESTIONNAIRE COLLECTION SYSTEM

32.   This section describes the software architecture features and the most interesting aspects of a software system that provides an Internet-accessible, easy to use online questionnaire to those wishing to respond via the Web.

A.   The software architecture of the electronic questionnaire

33.   This application aims to provide the households involved in the pilot census with an online questionnaire, as an alternative to the traditional paper questionnaire.

B.   Application architecture

34.   The following diagram summarises the application architecture.

Figure 2. The system software architecture
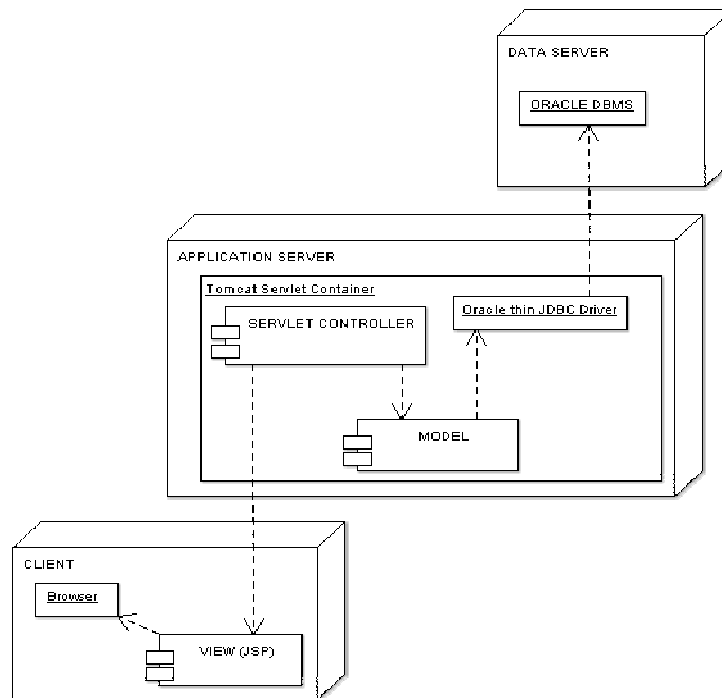
35.   The *application server* is the central architecture module, as it provides the infrastructure and support functions needed for the application execution.

36.   The *data server* provides the infrastructure and support functions needed for database access through network protocols. In this system the data server is an Oracle database.

37.   The *client* connects to the application server using any web browser, to access the available application services.

38.   The architectural model used to design the web application is the Model-View-Controller (MVC) pattern. It divides the application into three distinct logical levels, which manage the application status, the data presentation logic and the transformation of user requests into actions executed by the Model level.

39.   This is the only choice possible when it comes to building a complex, modular and easy-maintenance system, as it guarantees:

    (a)   Flexibility: the application is divided into several components based on their role; this allows new actions or views to be added without having to rewrite the entire application;

    (b)   Reuse: components can be easily integrated into other applications;

    (c)   Scalability: component separation allows for easy addition of new components when necessary;

    (d)   Security: central action management makes it easier to control access to actions and data.

C.   Implementation

40.   The web application was developed using the Java Enterprise Edition (version 5) language, which provides *servlet* and *Java Server Pages* (JSP) technologies.

41.   Servlets are software modules extending web server functions for http requests needing particular processing. JSP technology allows dynamic generation of HTML pages customised to user requests or application status.

42.   The application framework was developed to obtain an application that can be easily reused in similar surveys. The framework is a modular software architecture consisting of several classes and interfaces offering basic functions for data collection and validation and database interaction.

D.   Online questionnaire features

43.   The collection system is accessible through an Internet address reported in the letter accompanying the paper questionnaire sent to households.

44.   Access credentials consist of an identifier and a password. One credential will be given in the accompanying letter, while the other is an identifier known by both Istat and the household in question (for example the tax code of the person to whom the questionnaire is sent). The system guides users through the questionnaire step by step with a message system describing the actions to be performed on each page.

45.   The questionnaire is completed progressively through the following sections:

    (a)   list of persons usually living in the dwelling ("A" list);

    (b)   coded description of the household structure;

    (c)   list of persons not usually living in the dwelling;

    (d)   general information on the household and the dwelling;

    (e)   an individual form for each person usually living in the dwelling.

46.   An "index page" provides the user with an overview of the completion status of each section and links the sections and individual forms.

47.   The individual form completion process faithfully follows that of the paper questionnaire in all cases. The completion path is automatically shaped by the question sequence and answers to filter questions. The electronic questionnaire has been divided into pages to be browsed on screen, in such a way as to minimise their number while maintaining system usability. It is always possible to save partial data on the system server, in case the user wants to pause the process. Partial data is also saved automatically each time a new page is accessed.

48.   The user interface makes use of standard web form graphical elements: text box, radio buttons, check box, etc. This conforms the system to the most common metaphors, with which web users should already be familiar.

49.   The completion path is guided by logical rules that enable or disable questions according to the provided answers. For example, questions on nursery school attendance are disabled for persons whose age (calculated from the provided birth date) is higher than six.

50.   Predefined rules also enable the system to detect missing information and flag it up as error messages when users try to save data. They cannot complete the rest of the questionnaire until they provide the missing answers. The same rule management mechanism detects discrepancies between users' answers and the expected values for the corresponding variables and flags them up as errors. For example, range checks are performed on day, month and year values in birth dates or on hour and minute values in time data.

51.   Final data upload is possible once all sections and individual forms have been filled in. With this action, users declare they have completed the questionnaire. Before final upload, a PDF summary of the answers provided can be stored on the user's computer and printed.

52.   Once the data have been successfully uploaded users are presented with a thank you page confirming data receipt; again this in PDF format for local storage and printout.

V.   CONCLUSIONS

53.   This paper described the main features of an integrated online system for multi-channel data collection and census management. The system will be used for a pilot census performed at the end of October 2009. System specifications derive almost directly from the methodological innovations of the Italian 2011 population census. The higher complexity of the collection process is aimed at reducing the number of enumerators, which is achieved through a more sophisticated work process for the municipal back office. This requires a more complex monitoring system, which becomes a field operation management system: it guides targeted enumerator actions in the census tracts and allows efficient monitoring of the multi-channel questionnaire collection process.

54.   The pilot census should provide useful feedback on both the usability of the online questionnaire for households choosing this return mode and the probable efficiency of municipal operators in using the census management system.

*****