Managing census complexity through highly integrated web systems[1]

Federico Benassi,[2] Mauro Bruno[3], Maura Giacummo,[4] Marco Silipo[5], Giulia Vaste,[6] Donatella Zindato[7],

Abstract

During 2011 General Population and Housing Census methodological innovations towards the planning of a register-based census were introduced. Such innovations allowed to reach a satisfactory cost-benefit balance but increased the survey complexity and the risk of errors.
Istat focus on web technologies was aimed at performing an innovative census, both in terms of methodology and costs and of data dissemination timeliness.
This paper shows how a highly integrated web information system was used to manage a census workflow involving several actors and multiple phases and integrating different data sources.
The success of 2011 Census led Istat to adopt such technological infrastructure as the core of the 'Continuous General Population and Housing Census'.

Keywords: register-supported census, multimode data collection, survey management system.

1. Introduction

The 2011 census round saw the introduction of important methodological and technical innovations in the Italian population census. Among the most important ones were: the use of municipality population registers (LAC) as enumeration lists; the questionnaires mail out; a multimode data collection system; the on line crosscheck of census data and population register's records; an enumeration strategy differentiated according to the size of the municipality. Such innovations were designed to reduce problems related to conventional censuses (operational burden on municipal census offices, long delays between data collection and data dissemination, respondents' burden) and represent a first step towards the planning of a register-based census.

[1] The authors of the article would like to thank Francesca Bruno, Ph.D. student at Cornell University, for her significant help revising the text. The views expressed in this paper are solely those of the authors and do not involve the responsibility of Istat.
[2] Ricercatore (Istat), e-mail: benassi@istat.it.
[3] Tecnologo (Istat), e-mail: mbruno@istat.it.
[4] Tecnologo (Istat), e-mail: magiacum@istat.it.
[5] Tecnologo (Istat), e-mail: silipo@istat.it.
[6] Tecnologo (Istat), e-mail: vaste@istat.it.
[7] Primo Tecnologo (Istat), e-mail: zindato@istat.it.

Moreover, the way to cope with new types of errors had to be found: errors due to the use of registers such as coverage errors or errors due to time misalignments; errors due to the new data collection techniques, which allow greater flexibility but also determine a greater possibility of enumeration duplications.

Namely, an answer to the following question had to be provided: how can we minimize any possible errors? Or, in other words, how can we perform a *quality* census, to be at the same time (i) innovative in methodology as well as able to improve efficiency both in costs and in data dissemination timeliness; (ii) easy to be managed by census operators; and (iii) monitored in all its complex phases?

The Italian Statistical Institute's answer to these issues was a census focused on web technologies. Istat developed a highly integrated web information system that supported all of the different phases of the enumeration process. Such system, becoming itself one of the main innovations of the new strategy, was composed of three web applications: a) a web based management and monitoring system (SGR) accessible to all census operators, supporting every enumeration process activity; b) the online questionnaire (QPOP); c) an online documentation system, including different kinds of contents such as wiki, faq, manuals, legislation, etc.

Aim of this paper is to describe the impact that methodological innovations had from the technology point of view, to present the character and the main features of the technological architecture implemented by Istat and to analyze in details SGR functional logic. This last represented in fact, in the authors views, the joining link between planning expectations and empirical problems inherent to the different phases of the census.

The paper is organized as follows: section 2 provides a description of the main features of the new census strategy and of the main functions of the IT system designed for managing it; section 3 outlines the system architecture, focusing on the technological innovations; section 4 describes in detail the main functionalities of SGR, pointing out how it supported the complex census process, allowing cooperative operators' work and data integration; section 5 presents some concluding remarks.

2. New methods and techniques for the 2011 Italian census

2.1 The 2001 experience and the need of a new strategy

A decennial population census has been taken in Italy since 1861, based on the conventional methodology of complete field enumeration (so called "door-to-door" enumeration). Census forms were delivered and collected by enumerators and self-filled in by respondents. All information was collected and processed on a complete basis (without making use of any sampling technique) while the same economic, human and organisational resources were allocated to every household.

Obvious main goals of the census are the determination of the legal population and the collection of information on its main demographic and socio-economic characteristics. Furthermore, there is a third main goal to be achieved by the Italian census: the update of Municipal Population Registers (*Anagrafi*) on the basis of census results, as provided by the *Law on Population Registers* (*Regolamento Anagrafico*). This entails performing a crosscheck of census data and population register's records (the so-called *confronto*

censimento-anagrafe) concurrently to field-enumeration (Mastroluca and Zindato 2009).

Some figures will help us to give an idea of what a complex and demanding organization was required by the conventional census. In 2001, 22 million private households, accounting for a total of more than 57 million persons, were enumerated by 100.000 enumerators and 10.000 co-ordinators, organised in a network of 8.101 Municipal Census Offices (MCO's) and 103 Provincial Census Offices (PCO's).

The two main actors of this huge operation were: Istat, who has the responsibility for designing and coordinating census activities and processing and disseminating census data; and Municipal Census Offices, who are entrusted the responsibility of fieldwork, of revising completed questionnaires and of crosschecking census data and municipality records (and who, in most cases, contribute to the census with their own financial resources to the state allocated budget).

As to the 2011 census, a number of factors raised questions about the appropriateness of continuing to rely on conventional methodology. Namely, among these were: a) the huge organizational effort imposed on municipalities, exposed to a sudden and time-concentrated increase of workload; b) the need of improving of data dissemination timeliness; c) the increasing difficulty by enumerators of finding people at home, due to changes in both population life-style and structure (e.g. growing percentage of one-person households or of the so-called *dink* - double income no kids - couples), especially in larger municipalities; d) an increasing feeling of both dislike towards the census and public concern for confidentiality.

In order to identify the main critical points of fieldwork organization, a number of studies were conducted (Fortini et. al. 2007), pointing out the following as the main critical issues of 2001 census:

a) need to establish, co-ordinate and upkeep a massive network of enumerators and coordinators, with high difficulties of finding adequately skilled resources, high turn over rates and subsequent training problems;
b) high number of actors, highly differing from one another as to size, capacities, resources;
c) strong delays concerning the collection phase (the delivery of questionnaires was begun on time only by the 28% of municipalities while a mere 2% of them ended the collection according to the schedule).

Furthermore, the organizational impact of census operations turned out to be strongly dependent on the municipality population size: information from the 2001 census monitoring system showed that the biggest municipalities had the largest difficulties in meeting the field operations' deadlines while small municipalities struggled to cope with financial problems (Fortini et. al. 2007).

The results of the Pilot Survey held in 2009, designed in order to test several alternative enumeration strategies (Cassata and Tamburrano 2011), further proved the need of a modular and flexible census strategy, aimed at minimizing the aforementioned criticalities and taking into account the demographical and sociological changes occurred in the Italian society since the 2001 census.

2.2 Main features of 2011 census strategy

Combining the study of census experiences of other countries with a more effective use of administrative data held by municipality population registers, a completely new strategy

was designed (Kotzamanis et al. 2004; Abbatini et al. 2007; Ferruzza et. al. 2007). Such a strategy relies on a number of methodological and technical innovations, and on the crucial role of a census web management system, being the backbone of every phase of the enumeration process (Istat 2009).

Main features of the 2011 census have continued to be the completeness and simultaneity of the population count, but the fieldwork was guided by registers and supported by the use of new data collection techniques and new territorial instruments designed to improve coverage and quality of the enumeration.

The "door-to-door" census became a *register-supported* census, implemented by means of questionnaires' *mail out* to households registered into municipal population registers. Namely, the 8092 Municipality Population Registers were used as lists of households (and addresses) to which census questionnaires were mailed out. Self-completed questionnaires were collected by a multimode system which included Internet, return at any post office in Italy, return to Municipal Collection Centres and, finally, targeted recovery of non-response by enumerators.

Thanks to the use of questionnaires *mail out* (instead of enumerators' delivery) and of a multimode data collection system (where enumerators are just one of the possible return modes, and hierarchically the last one), the front-office staff recorded a dramatic reduction (about 40%) and a great flexibility was allowed to respondents (Picci and Sindoni 2012).

A major issue concerning the new strategy was represented by over-coverage and undercoverage list errors typically affecting a *register-supported* enumeration (i.e. a number of households included in the municipality population register might be no longer residing in the municipality and, conversely, a number of households actually residing on the municipality territory might not be included in the population register). While over-coverage was 'automatically' corrected by a field enumeration that relied on questionnaire *mail out* to units included in the list (by assuming that households residing no more in the municipality would not receive therefore not return the questionnaire), specific measures were required in order to manage potential undercoverage. To this aim, data provided by different sources (such as the revenues agency or foreigners permits to stay) were used to set up a list of persons not included in registers but potentially residing in the municipality. An additional list was based on the pre-census Address Numbers' Survey, containing information on potentially inhabited housing units for which there was no corresponding entry in the municipality records. Enumerators have been sent out to look for and deliver questionnaires at these addresses. Another basic feature of the new strategy was its modular nature. The necessity to differentiate census organization according to the needs and capacities of the different actors required a strategy consisting of a set of modules to be applied flexibly according to the size of the municipality. More precisely, municipalities were divided in two main size categories, and a different combination of modules was planned for each of them (Zindato 2012) (see Table 1).

A major change concerning only largest municipalities (i.e. those with at least 20.000 inhabitants and all province seats) was the shift towards the production of estimations concerning the socio-economic set of census variables. These estimations were produced by using a *long form* on a sample of households. Data produced in this way are significant at a census area (grouping of contiguous and homogeneous enumeration areas) level.

In the same subset of municipalities, a pre-census addresses' survey was carried out to the aim of producing a field-checked geo-coded list of addresses with the related number of housing units, in order to produce auxiliary information to be used to limit undercounting. The additional costs required by the setting up of an address list for the smallest municipalities compared to the corresponding advantages in terms of accuracy and quality of the count restricted the planning of this operation to the largest municipalities (Picci and Sindoni 2012).

Table 1 - Census strategy main modules by municipality type

MODULE	Municipality type	
	20.000 inhabitants or more	<20.000 inhabitants
Pre-census Address Numbers' Survey	X	
Setting up of census areas	X	
Use of pre-census lists derived from Municipality Population Registers	X	X
Use of sampling for collecting socio-economic information (short form/long form strategy)	X	
Crosscheck of census data and population register's records	X	X

Finally, a last but much-important change concerning all municipalities was the actual performing of the crosscheck of census data and population register's records at the same time of the enumeration and through a standardized and public (i.e. visible to Istat) instrument.

2.3 The new strategy adopted and IT management required

The strategy designed for the 2011 census aimed at reducing both municipalities workload and respondents' burden, at holding down costs and at enhancing data dissemination timeliness. The basic idea was to differentiate organization according to the needs and capacities of the different field-work actors and to reduce the burden on respondents by giving them the possibility to choose the return mode that would better satisfy their needs. The three main pillars of such a strategy were as follows:

a) lesser use of front office staff but reinforcement of municipalities back office activities;
b) flexibility of fieldwork organization within Municipal Census Offices;
c) flexibility of collection techniques.

Indeed, the management of a modular and flexible strategy represented a big challenge. On one hand it helped solving problems that traditionally had a great impact on census process, negatively affecting the timeliness of data dissemination. On the other hand, the introduction of this new strategy implied a higher level of complexity and a multiplication of risk factors (Benassi et. al. 2013).

The management of such a complex and diverse enumeration strategy entailed the need of a very flexible web management system. As already mentioned, such a web management system was in fact crucial to the performing and success of the entire census, being a

complete instrument that guided and supported census operators during all the survey phases. It replaced previous monitoring systems, which offered a dashboard with quantitative indicators for tracking questionnaires without providing any management or support tools to census operators (Istat 2012). Basically, it was designed to provide the different users of the system with: (i) up-to-date information at different aggregation levels, including single questionnaire level; (ii) a tool for cooperative working, guided through a forced workflow of questionnaire life-cycle.

The coexistence of different return modes i.e. of information coming from different sources (on-line questionnaire, Post Offices monitoring system, MCOs) required a web IT system constantly updated enabling census staff to follow the status of every questionnaire over time.

The system, accessible online to all of the different levels of census staff, enabled the status of every individual questionnaire to be followed in almost real time, thus allowing the targeted recovery of missing questionnaires: the availability of constantly updated information on the status of each questionnaire enabled enumerators to be directed only to households to which the questionnaire was sent but not yet returned (crosscheck of census and municipality records)**.**

As already mentioned, auxiliary lists enabled the targeted and systematic recovery of individuals not registered in municipal records. Such lists were integrated and loaded into the system, thus allowing enumerators to systematically check undercoverage, as long as they were on the field for the recovery of non-response.

Furthermore, the system was designed to automate back-office work and to guarantee flexibility to fieldwork organization within each Municipal Census Office. Municipal Census Offices managers had to assign an organisational role and a system profile to every user and allocate enumeration areas to enumerators. Each census office could thus freely decide how to distribute work in terms of assignment of enumeration areas to enumerators and back office work to operators. A hierarchical organisation could also be defined by setting dependency relationships between staff with a coordinator role and other staff and of enumerators to co-ordinators.

The system also included an important function to be used for performing the on-line crosscheck between census data and population registers and for the production of the related accounts. By entering in SGR the identification data of enumerated people, as long as questionnaires were returned, and comparing them with municipality records updated to 8 October 2011 (available in the system) Municipality officers performed a real-time crosscheck which enabled the earlier conclusion of the census and the earlier dissemination of the results. In fact, the first and very important final data (i.e. data on the municipalities' legal population by sex, age and citizenship) were timely released before the end of questionnaires data capture, on the basis of data entered in SGR.

Finally, being as well a monitoring system, SGR also allowed to produce census progress reports.

The census web based management and monitoring system was part of a general strategy aiming at minimising errors, reducing organizational workload and holding down costs. The other fundamental components of the overall strategy were the online questionnaire (QPOP) and the online documentation system.

The broad use of QPOP (33,4% of returned questionnaires) resulted not only in a significant reduction of municipalities front-office work (reduction in the number of

enumerators) but also of back-office work in that the quality revision to be performed by MCOs on paper questionnaires need not be done on electronic questionnaires (Picci and Sindoni 2012). In fact, QPOP guided respondents in the correct compilation of their questionnaire through consistency rules and error checking. Moreover, QPOP presented to the respondent only the correct set and sequence of questions to be filled in, so that the online compilation turned out to be easier, faster and less error prone than the paper form (Virgillito and Tininini 2012). The reduction in the number of pages scanned also resulted in some costs' reduction, even though a remarkable reduction could have been achieved only by avoiding sending all paper questionnaires.

Finally, a not negligible role was played by the online documentation system, which functioned as the reference site for the network of SGR operators, who used it to access documentation materials, such as manuals and legislative references, but also as an up-to-date information site for contents such as wiki, faq and communications. It has to be underlined the importance for such a complex and vast organization of an open space for sharing standardized information, contributing to reduce problems due to informational asymmetry.

3. System architecture

One of the main complexities of setting up such an important and business critical system is the integration of multiple data sources and the ability to guarantee the availability of the whole system on a 24/7 basis, in particular during peak hours.

These goals can be achieved only by designing and implementing a flexible architecture, which is based on balanced and replicated systems and uses a consolidated middleware infrastructure where components can be plugged in easily and without massive code reworks.

As shown in Figure 1, SGR was a core part of such architecture, integrating its modules with the questionnaire application that was used by respondents for data entry but also from MCOs who chose to enter data from the paper questionnaires.

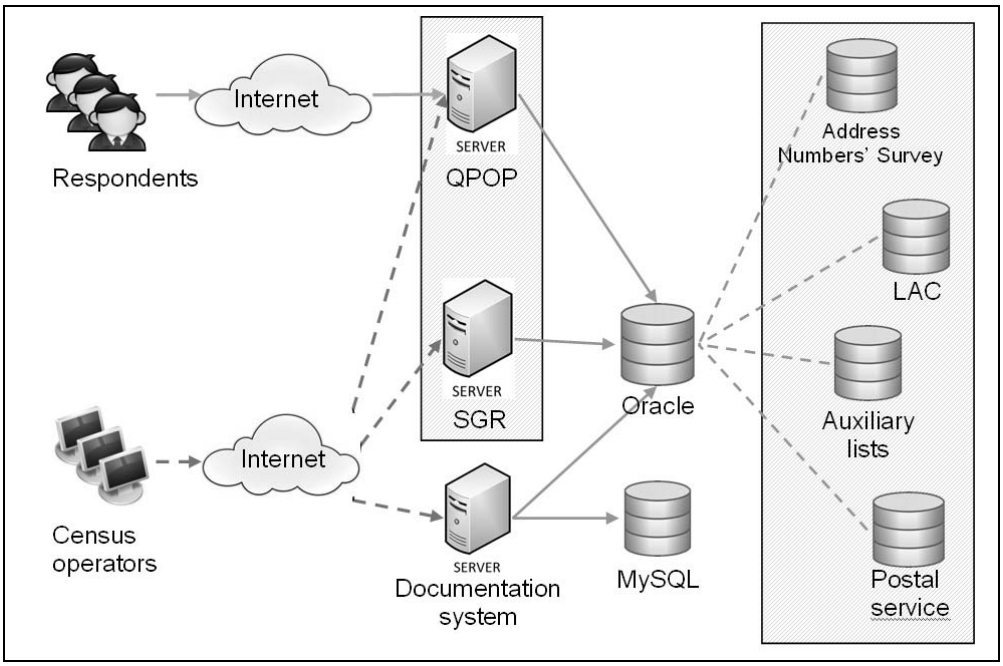
Data on which SGR was based relied on a very complex ORACLE database which had to be accurately tuned for data access optimization from a very high amount of users, including respondents and operators.

3.1 Software technologies

The core software infrastructure of SGR and QPOP followed the Model-View-Controller design pattern. The main technology on which SGR was based is the Java Enterprise Edition (JEE) platform. Java was the core language used for the development of the last Agriculture Census management system. The team that was in charge of the development of SGR for the Population census took advantage of the previous experience, improving the software in terms of security and general quality, reusing many parts of the infrastructure but switching to the state-of-the-art versions of the frameworks chosen as building blocks. The heavy use of frameworks proved to provide a significantly positive impact, producing a cleaner code that was easy to write, test and maintain, resulting in a more robust application. The implementation exploited in particular three widespread open source frameworks: Struts2, Spring and Hibernate.

Hibernate proved to be fundamental for developers, as it acts as a software layer which eases working with database tables through a simple model based on Java objects, called "beans". This speeds up development time and delegates the SQL interaction with the DBMS to the framework: programmers only have to deal with the generally more familiar Java syntax.

Figure 1 – System architecture



Another technology widely used in SGR was AJAX: this Javascript-based technology acts as glue between the GUI and the server-side components of the system, allowing a tight interaction among the controls available on the end-user interface (buttons, lists, collapsible data sections) and consequent real-time actions happening on the server. This technology brought real improvements to the end-user, who was actually guided in his work since the data SGR presented to him were always fetched from the database using small and focused queries. This also brought important performance benefits: for example, when the user selected a province from a dropdown list and was called to select a municipality to work with, the municipalities he could choose from were a subset which depended on the previous choice of the province.

Another component of the architecture was the online documentation portal developed by means of a widely popular open source CMS (Typo3), entirely written in PHP. This led to an additional integration problem due to the fact that Typo3's database management system is MySQL while SGR data was Oracle-based: since any SGR user had to be authenticated on both systems, user data had to be constantly synchronized, through a custom script.

The components of the system architecture are shown in Figure 1. QPOP, SGR and the Documentation system were three integrated web applications. The integration mainly concerned the authentication system and the management of questionnaires returns. With respect to the authentication system, on the one hand, a single-sign-on mechanism was provided for SGR and QPOP, allowing an SGR operator to access QPOP for the data entry of his assigned questionnaires. On the other hand, SGR and the Documentation system authentication tables were synchronized allowing users to access both applications with the same credentials. Concerning the questionnaire returns, QPOP and SGR shared information on the questionnaire status: when a citizen completed the online compilation of his questionnaire this information was available in SGR; when a questionnaire was returned and registered in SGR the online compilation was disabled.

The database was loaded with normalized data from different sources, e.g. Address Numbers' Survey database, LAC, auxiliary lists, Postal service. A detailed description of the integration of such sources is provided in Section 4.

3.2 Security issues

Security is the major concern in a system where the number of users involved reaches a potential of many millions. When the lights of the media focus on such an event like the population census, malicious and talented programmers in search of popularity often undertake a new challenge with themselves. Thus, the organization of the defence has to be really accurate and must take into account most of the state-of-the-art security techniques (OWASP 2013, SAFECode 2013).

The first mechanisms that must be put in place are the most widely consolidated security best practices: authentication based system, strong encryption algorithms, role-based authorization levels and strong user-profiled URL protection.

Since SGR and QPOP relied on a cluster of Tomcat Application Servers, it was straightforward to exploit the capabilities of this middleware infrastructure: in order to implement Authentication using Single Sign On, Tomcat realms and the *j_security_check* mechanisms were used, storing passwords as SHA-256 hashes on the systems. As a consequence a breach of security would not have disclosed users' credentials but only their encrypted representation.

Authorization was implemented using the core Java Enterprise Edition patterns i.e. a complex deployment descriptor with a fine-grained set of operator profiles that came out from a territorial and functional perspective. Any SGR or QPOP user, once logged in the system, could navigate only between the set of functionalities of his competence (through the JEE security-constraints and their correspondent auth-constraints) and could browse only data related to his territory or personally assigned to him (questionnaires in particular).

This last defence against the so-called "Privilege Escalation attack" in particular was achieved through a further check on the backend of SGR and QPOP. Suitable business logic was implemented in order to verify the user authorization to access the requested information.

3.3 Testing phase

The system underwent both functional and infrastructural tests (SWEBOK 2004, OWASP 2013). Functional tests were performed by an internal team in order to verify the

consistency against the software requirements. An additional set of tests on the intermediate releases of the system were demanded to an external team of testers chosen among the census staff across the national territory.

The infrastructural test phases involved a partner company that was asked to perform several vulnerability assessments against SGR and QPOP in order to evaluate the stability and the security of the applications. Tests were performed both using the Black Box approach (that is, without knowing any detail of the systems to be violated) and the White Box approach (using valid credentials and then trying to get more information than allowed to your user's profile). The analysis was split in three phases:

- Vulnerability Assessment: the system was attacked from an external user trying to access protected data exploiting application vulnerabilities by means of techniques like Information Gathering, Data Tampering, Buffer Overflow, XSS and SQL injection. In particular the Data Tampering technique was used in order to check the system against the risk of Privilege Escalation (authenticated users that can increase their authorization level using particular tools);
- Code Review: using static analysis tools it was possible to determine if the applications were compliant to coding best-practices, if they were subject to memory leaks or DB connection leaks, if there was unreached or unused code or if there were code snippets that could cause performance decrements;
- Penetration Test: the system infrastructure was tested against Denial of Service attacks and a port scanning was performed.

All tests were conducted on both a specific test environment and the final production infrastructure, to make sure results of the test were consistent and reliable. The final results proved that both SGR and QPOP were not vulnerable to attacks, and the success of the overall infrastructure during the survey confirmed its quality.

4 The web based management and monitoring system

SGR was designed as a collection of functions, each related to a process phase and customized according to the user profile that accesses it. About eighty functions, grouped in menus and sub-menus, were developed to support all the census activities. Therefore SGR resulted in a modular, flexible, and scalable system which allowed an agile development process. Although more than twenty developers were involved in the realization of SGR, the plug-and-play design of SGR allowed a strong cooperation and a rapid development of all the functionalities.

The system was dynamically customized according to the profile of the logged user: both the menus and sub-menus were personalized displaying only the functions the operator was authorized to use.

To illustrate the functionalities offered by SGR we focus on the following macro-areas: a) integration of multiple data sources (registers, Address Numbers' Survey, etc.); b) interaction of different actors of the survey process (enumerators, supervisors, Istat personnel, postal service); c) implementation of the workflow for the questionnaire life-cycle management; and d) up-to-date monitoring of the survey progress.

4.1 Integration of multiple data sources

As we said before, an important innovation of the 2011 census was the use of municipality population registers (LAC): questionnaires were personalized with information concerning the householder and mailed out by the Italian postal service. In particular, the SGR database was loaded with normalized data, almost sixty millions of individuals, coming from LAC dating to 31/12/2010. Further, to take into account population flows in the period between 31/12/2010 and the census date, i.e. 9/10/2011, a second data loading was performed for municipalities with more than twenty thousands respondents. Small municipalities were supported in such operations by SGR through suitable functionalities.

As already mentioned, LAC can be affected by coverage errors. So, data provided by other sources, such as auxiliary lists and Address Numbers' Survey, were integrated in the system to help detect undercoverage. In particular, the Address Numbers' Survey allowed the detection of buildings with no corresponding LAC individuals: almost nine millions of possible undercoverage signals were loaded in the system and checked by the enumerators.

SGR managed also the different questionnaire return modes offered to respondents, i.e. web compilation, return to postal office, to Municipal Collection Centers and to enumerators. On the one hand the integration between QPOP and SGR allowed a real-time monitoring of the online compilations; on the other hand the integration with services provided by the Italian postal service was necessary in order to load information about both deliveries of the personalized questionnaires to the respondents and returns to the postal offices.

4.2 Interaction of different actors of the enumeration process

SGR included several user profiles, each characterized by a different territorial visibility (national, regional, provincial, municipal) and a number of available functions (Table 2).

User profiles with national visibility had monitoring functionalities that enabled them to monitor the survey progress on the whole territory and to take strategic decisions during the survey. User profiles with regional visibility were provided to Istat regional census staff. The regional employees' duty was to organise and coordinate the survey on their territory and to support the municipal operators. The municipal operator was the key profile in SGR, being in fact responsible for all operative phases of the census. Such users could: (i) define the local survey network, i.e. supervisors and related enumerators; (ii) assign enumeration areas with related questionnaires to the enumerators and (iii) monitor the progress in the questionnaires life-cycle. Enumerators and supervisors carried out field work and back-office activities, such as registration of the questionnaire returns to the Municipal Collection Centers, check of data provided by the Address Numbers' Survey in order to detect undercoverage, etc.

Actually, SGR gave the possibility to create autonomously the survey network and more than eighty five thousand operators' accounts were created. This was a significant result since the creation of the survey network was not managed as a centralized task but it was distributed on the whole territory.

Table 2 - User profiles

USER PROFILE	Functionalities	Territorial visibility
Istat	Monitoring functionalities at all territorial level, i.e. national, regional, provincial, municipal and enumeration areas level	National
URC	Monitoring and support functionalities at regional, provincial, municipal, and enumeration areas level	Regional
UPC	Monitoring and support functionalities at provincial, municipal, and enumeration areas level	Provincial
UCC	Monitoring and support functionalities at municipal and enumeration areas level. Functionalities for creating the municipal operators network. Functionalities for assigning, coordinating and supervising both enumerators and supervisors work. Operative functionalities for field work and back-office activities.	Municipal
CoC	Functionalities for assigning, coordinating and supervising enumerators work. Operative functionalities for field work and back-office activities.	Assigned enumeration areas
Ril	Operative functionalities for field work and back-office activities	Assigned enumeration areas

4.3 Implementation of the workflow for the questionnaire life-cycle management

SGR guided the operators to conduct the survey correctly, offering a fixed path through the questionnaire working phases. Each questionnaire working phase was linked to a state. The transition between two different states was realized through SGR functions or through external operations, such as the completion of the online compilation. Each function had both pre-condition states, i.e. states that allowed the function use, and post-condition states, i.e. states assigned by the function to the questionnaire. In this way SGR defined a flow of questionnaire states, which guided and forced the questionnaire life-cycle. Such questionnaire life-cycle management allowed also cooperative operators work. For example we can consider the following scenario: 1) a back-office operator registers the paper questionnaire return; 2) SGR updates the questionnaire state; 3) consequently the enumerator, responsible for that questionnaire management, is informed of the return and can proceed with following working phases.

Thus SGR was a distributed workflow system, in which, on the one hand, each operator worked autonomously and, on the other, a centralized monitoring of the overall census activities was provided. As a result, using SGR as a survey tool allowed for cost-effectiveness, real-time management, support for cooperative work and on-going monitoring.

The most important function in SGR was the “diary”: a control panel that showed to the survey operator an up-to-date list of his assigned questionnaires, built by the different sources described in section 4.1. Each element of the list displayed the respondent’s name and address, the questionnaire state, and the operations already performed on the questionnaire. The operator could modify the questionnaire state: on the basis of the current questionnaire state the diary showed the possible next states. Further the diary displayed the

information to be edited according to the state transition. In such a way the diary allowed to manage the significant aspects of the process, such as the coexistence of different questionnaire returns offered to respondents. Each return put the questionnaire in a defined state, through internal or external functions, i.e. web compilation, loading of postal office information in the system database and census operator registration. Such information, which was visible in the diary, allowed the complete monitoring of the returns. As a result, the enumerators were able to check only the respondents that had not returned the questionnaire. The enumerator's work was thus more efficient and the quality of the process increased.

Through the diary it was also possible to enter information concerning questionnaire summary data. Such information was mandatory in order to reach a questionnaire final state. The availability of summary data in SGR has been a key element for a rapid dissemination of provisional data.

The diary resulted in a complete instrument for all the operators involved in the survey process: the enumerators used it as a control panel that showed them an up-to-date list of assigned questionnaires and guided their field-work; the supervisors used it as an instrument to monitor the work of their assigned enumerators; the municipal operators referred to the diary to monitor the survey progress in enumeration areas.

4.4 Up-to-date monitoring of the enumeration progress

SGR allowed a comprehensive monitoring of the survey process. On the one hand, SGR offered a detailed supervising of all the operations performed on each questionnaire, displaying author and date of each operation. On the other hand, several reports were provided in order to show the progress of each survey phase: delivery of the questionnaires by the postal service; returns of the questionnaires; survey progress on the basis of questionnaire states; activity of the operators (with a high level of detail).

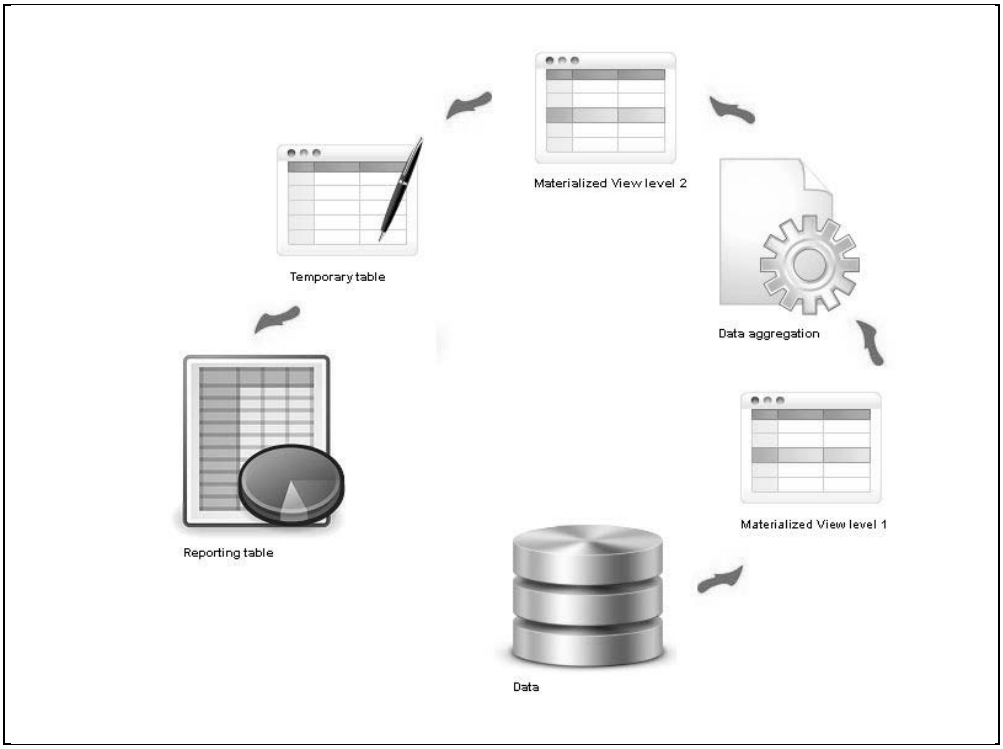
In order to provide a continuous monitoring, without affecting the performance of the application, the reports were updated by the system at regular intervals of one hour. All upgrade procedures were handled within the database, delegating to the application server only the visualization of already executed summaries.

In the database were created materialized views and temporary tables that, with different levels of territorial aggregation, produced the required results (Figure 2). The materialized views, being updated incrementally, allowed saving only those variations that occurred over a period of time, with significant savings in execution times.

Reports were available at different data aggregation level. In particular, it was possible to display data concerning a given operator or data at different territorial levels, i.e. national, regional, provincial, and municipal.

Every report could be exported in the most common formats like XLS and PDF, but also in a CSV plain text file, allowing users to import data in any custom tool.

Figure 2 – Monitoring reports production



5. Concluding remarks

SGR turned out to be a very useful instrument that supported the 2011 Italian population census: it was daily accessed by more than one hundred thousands users during the survey; it supported a complex mixed-mode collection system, with about 8.500.000 online questionnaires (33,4% of the total returns), 7.826.000 questionnaires returned at Municipal Collection Centers, over three million questionnaires returned to enumerators and the remaining 20% returned to post offices.

Even though measurements on the quality of the survey results have not been performed yet, the management of the survey pointed out significant improvements towards a cost-effective and *quality* census. SGR greatly contributed to achieving such results.

First, it supported a register-supported census, allowing a reduction in the related errors. Such results were achieved through the use of multiple data sources as well as a constant and complete monitoring of all survey phases.

Second, SGR offered a predefined and forced workflow in the questionnaire life-cycle management; this enables enumerators' work to be less prone to errors and more cooperative.

Third, SGR was used by operators with different responsibilities, providing suitable

functionalities for each profile: (i) the “diary” effectively supported the enumerators’ field work and (ii) reports allowed monitoring to users with different territorial visibility, also supporting the strategic decisions of Istat census managers.

SGR was originally adopted in the 2010 Agricultural Census, and it was recently used in the 2011 Industry and Services Census. In this last survey, SGR was enriched with new functionalities, such as micro data check on the base of predefined rules. Due to SGR the data production process was agile and efficient.

Since SGR software architecture is mainly framework-based and standard compliant, it revealed itself much suitable for constituting the foundation of a set of additional web systems that have been recently implemented in order to support the subsequent stages of the census process. As few but significant examples worth mentioning are the Post Enumeration Survey (PES) and the System for the Review of the Municipal Population Registers (Sirea).

Istat technological innovation plans include an SGR generalization with the aim to adopt it as a system for the management of many other surveys. In particular, it will be a central component in the design of the ‘Continuous Population Census’, which is due to replace the ‘one shot’ census, with its first cycle beginning on 2016.

References

Abbatini D., L. Cassata, F. Martire, A. Reale, G. Ruocco and D. Zindato. 2007. *La progettazione dei Censimenti Generali 2010-2011. 2 – Analisi comparativa di esperienze censuarie estere e valutazione di applicabilità di metodi e tecniche ai censimenti italiani.* Documenti ISTAT, n. 9. http://www.istat.it/it/files/2011/04/2007_9.pdf (April 2013).

Benassi F., L. Cassata, G. Sindoni and D. Zindato. 2013. "Tales from the 2011 Italian Population Census. The use of a multi-mode data collection system: lessons learnt and future challenges". Relazione presentata alla Conferenza: 5th Conference of the European Survey Research Association (ESRA). Ljubljana 15-19 July.

Cassata, L. and M.T. Tamburrano. 2011. *The 15th Population Census Pilot Survey: how the register driven census changes the enumerators role* in: S. Migani and M. Costa (Eds.) "Statistics in the 150 years from Italian unification", Serie Ricerche n. 10, Università di Bologna, Dipartimento di Scienze Statistiche "Paolo Fortunati", Bologna. http://amsacta.unibo.it/3202/1/Quaderni_2011_10_SIS2011_BookofShortPaper.pdf (April 2013).

Fortini M., G. Gallo, E. Paluzzi, A. Reale and A. Silvestrini. 2007. *La progettazione dei Censimenti Generali 2010-2012. 3 – Criticità di processo e di prodotto nel 14° Censimento generale della popolazione e delle abitazioni: aspetti rilevanti per la progettazione del 15° Censimento*. Documenti ISTAT, n. 10. http://www.istat.it/it/files/2011/04/2007_10.pdf (April 2013).

Ferruzza A., S. Mastroluca and D. Zindato. 2007. "I censimenti esteri: modelli a confronto alla luce dei regolamenti internazionali". Relazione presentata alla Conferenza: Censimenti generali 2010-2011. Criticità e innovazioni. Roma 21-22 November. http://www3.istat.it/istat/eventi/2007/interconferenza/interventi/Ferruzza_Mastroluca_Zindato.pdf (July 2013).

Istat. 2009. *A new strategy for the 2011 Italian Population Census. Product innovations and the compliance with CES Recommendations,* UNECE/CES Group of Experts on Population and Housing Censuses, Twelfth Meeting, Geneva, 28-30 October. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2009/5.e.pdf (April 2013).

Istat. 2012. *A new strategy for the 2011 Lessons learned from use of registers and geocoded databases in population and housing census,* Sixtieth plenary session, Paris, 6-8 June. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2012/22-IP_Italy.pdf (April 2013).

Kotzamanis B., G. Cantisani, A. Dekker, D. Logiadu Didika, M. N. Duquenne and a. Castori. 2004. *Documentation of the 2000 Round of Population and Housing Census in the EU, EFTA and Candidates Countries*, Luxemburg: Office for Official Publications of the European Communities. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-CC-04-003/EN/KS-CC-04-003-EN.PDF (April 2013).

Mastroluca, S. and D. Zindato. 2009. *Censimento della popolazione e delle Abitazioni*, in: Egidi, V. and A. Ferruzza (eds.) Navigando tra le fonti demografiche e sociali, Istituto Nazionale di Statistica, Roma. http://www3.istat.it/dati/catalogo/20100325_01/Navigando_tra_le_fonti_demografiche_sociali.pdf (July, 2013).

OWASP, Open Web Application Security Project https://www.owasp.org/index.php/Main_Page (July 2013)

Picci, M. and Sindoni G. 2012. "Nuovi metodi e tecniche per il censimento della popolazione. I primi numeri del Censimento 2011". Relazione presentata alla Conferenza: Interrogare le fonti 2: un confronto interdisciplinare sull'uso delle fonti statistiche. Napoli, 14-15 Giugno 2012.

SAFECode, *Fundamental Practices for Secure Software Development* http://www.safecode.org/publications/SAFECode_Dev_Practices0211.pdf (July 2013)

SWEBOK, *Guide to the Software Engineering Body of Knowledge 2004 version*, chapter 5. http://www.computer.org/portal/web/swebok/htmlformat (July 2013)

Virgillito, A. and L. Tininini. 2012. *The Web-based Data Collection in the Italian Population and Housing Census Meeting on the Management of Statistical Information Systems*. MSIS 2012. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2012/18_Italy.pdf (April 2013).

Zindato D. 2012. "Nuove prospettive per il Censimento della popolazione. Verso il censimento continuo". Relazione presentata alla Conferenza: Interrogare le fonti 2: un confronto interdisciplinare sull'uso delle fonti statistiche. Napoli, 14-15 Giugno 2012.