CONTRIBUTI ISTAT

n. 11/2008

Experimenting Data Capturing Techniques for Water Statistics

G. Di Bella e *S. Macchia*

Istat

SISTEMA STATISTICO NAZIONALE
ISTITUTO NAZIONALE DI STATISTICA

Le collane esistenti presso l'ISTAT - *Rivista di Statistica Ufficiale*, *Contributi ISTAT* e *Documenti ISTAT* - costituiscono strumenti per promuovere e valorizzare l'attività di ricerca e per diffondere i risultati degli studi svolti, in materia di statistica ufficiale, all'interno dell' ISTAT, del SISTAN, o da studiosi esterni.

La *Rivista di Statistica Ufficiale* accoglie lavori che hanno come oggetto la misurazione dei fenomeni economici, sociali, demografici e ambientali, la costruzione di sistemi informativi e di indicatori, le questioni di natura metodologica, tecnologica o istituzionale connesse al funzionamento dei sistemi statistici e al perseguimento dei fini della statistica ufficiale.

I lavori pubblicati in *Contributi ISTAT* sono diffusi allo scopo di stimolare il dibattito intorno ai risultati preliminari di ricerca in corso.

I *Documenti ISTAT* forniscono indicazioni su linee, progressi e miglioramenti di prodotto e di processo che caratterizzano l'attività dell'Istituto.

Il Comitato di redazione esamina le proposte di lavori da pubblicare nelle tre collane sopra indicate. Quelli pubblicati nei Contributi ISTAT e nei Documenti ISTAT sono valutati preventivamente dai dirigenti dell'Istituto, mentre i lavori pubblicati nella Rivista di Statistica Ufficiale sono subordinati al giudizio di referee esterni.

ISTAT – Produzione libraria e centro stampa

CONTRIBUTI ISTAT

n. 11/2008

Experimenting Data Capturing Techniques for Water Statistics

G. Di Bella(*) e *S. Macchia(**)*

(*) ISTAT - Servizio Acquisizione gestione e diffusione delle fonti amministrative
(**) ISTAT - Servizio Metodologie, tecnologie e software per la produzione dell'informazione Statistica

Contributi e Documenti Istat 2008

Istituto Nazionale di Statistica
Servizio Produzione Editoriale

Produzione libraria e centro stampa:
Carla Pecorario
Via Tuscolana, 1788 - 00173 Roma

Abstract: Official water statistics demand is increasing in the last years in order to better manage water resources and to guarantee the ecological protection of water bodies. Standard procedures and statistical methodologies to improve data quality are emerging within the several distributed national data producers. Regarding data collection through a questionnaire, in the last few years new procedures has been developed from the computer science point of view; in particular the use of the electronic questionnaire has offered a great number of tools also for the environment statistical surveys. In this paper the experience conducted in the Italian statistical office in surveying water services data using an electronic questionnaire is described.

Keywords: Water statistics, Water services, Water management, Statistical survey, Computer Assisted Interviewing

---

1. Introduction

Water statistics, as being part of a theoretical national water information system, cover a great number of fields considering they regard water both as a natural resource and as environmental media. Data concern water resources, water use, wastewater treatment, emissions, water quality, aquatic ecosystem analysis, water services management and policies. In Italy many organisms are involved in data collection, and many activities have been carried out in the last years around this issue, but the actual data availability and its quality do not meet the current demand. Water statistics production covering the national geographical level is still an open challenge.
Unlike water quality monitoring data (data for physical, chemical, biological and microbiological waters characterization) that are mainly treated by the Environmental Agencies, national environment statistics related to population, public administration and enterprises (generally considered as environmental pressure data) are in most cases surveyed by the National Statistical Institutes (NSI's). In this case statistical offices have the advantage of their traditional experience in collecting data through surveys using questionnaires submitted to a well-defined responding units set.
Regarding water statistics, that are a basic element of environmental information, Italian NSI (Istat) is involved in data collection about water services.
In this paper the Istat experience is described: in paragraph 2 the Oecd/Eurostat 'State of the Environment' Questionnaire is illustrated as water statistics demand. The paragraph 3 and 4 deal with the Istat project on water statistics collection. Then the data capturing procedure recently used and the advantages deriving from it are pointed out.

2. The Eurostat/Oecd water data framework

From an institutional point of view the main international data collection involving NSI's is the Joint Questionnaire on the State of the Environment - section Inland Waters (JQ –IW) managed by OECD (Organisation for Economic Co-operation and Development) and Eurostat (Statistical Office of the European Communities). The Questionnaire is the result of more then 20 years of experience inside OECD, Eurostat and participating member states (Amand and Montgomery, 2001).
The main objective of this Questionnaire is to collect, but also to harmonise, national environmental data. Subjects dealt with in the Inland Water section concern:

- water resources abstraction
- water resources use
- wastewater treatment
- production and disposal of sewage sludge
- discharge of wastewater into the environment
- quality of rivers and lakes

For water abstraction the complete framework proposed by the JQ-IW is reorganized in the Table 1.
For each sector operating abstraction, data refer to the amount of water abstracted (thousands m^3/y) from *Fresh surface water*, *Fresh ground water* and *Marine and brackish water*; data on *Desalinated water* and *Reused water* are also requested.
Abstraction sectors are different from one another in using distinct infrastructure networks: pumps, tanks, aqueducts.
The *Public water supply* regards water intended for human consumption, that is drinkable or that it is treated to reach the standard predefined quality. In the JQ-IW it is defined as 'Water supplied by economic units engaged in collection, purification and distribution of water (including desalting of sea water to produce water as the principal product of interest, and excluding system operation for agricultural purposes and treatment'. It may be also denoted as Water supply industry, coded 41 according to the International Standard Industrial Classification of all economic activities (ISIC Codes).

Table 1. JQ-IW framework for Annual water abstraction by source and by sector

ABSTRACTION SECTORS	SOURCES			TOTAL	Desalinated water	Reused water
	Fresh surface water	Fresh ground water	Marine and brackish water			
Public water supply			(shaded)			(shaded)
Agriculture, forestry, fishing						
of which irrigation						
Manufacturing industry						
of which industry-cooling						
Production of electricity						
Other activities						
Households			(shaded)			
TOTAL						

Data about abstraction of *Marine and brackish water* by the *Public water supply* is requested as *Desalinated water*, that is a necessary purification treatment for drinking water.
Other sectors are considered as self supply operators and in this case, depending on the specific use, water abstracted and supplied can be potable or not. For water use the JQ-IW framework can be summarized as in Table 2.

Table 2. JQ-IW framework for Annual water use by supply category and sector

SUPPLY CATEGORIES	USE SECTORS									
	Agriculture, forestry, fishing		All industrial activities			Production and distribution of electricity		Domestic sector		
		of which irrigation purposes		of which Total manufacturing industry	of which for cooling purposes		of which for cooling purposes		of which households	of which other activities
Public water supply		(shaded)								
Self supply										
TOTAL										

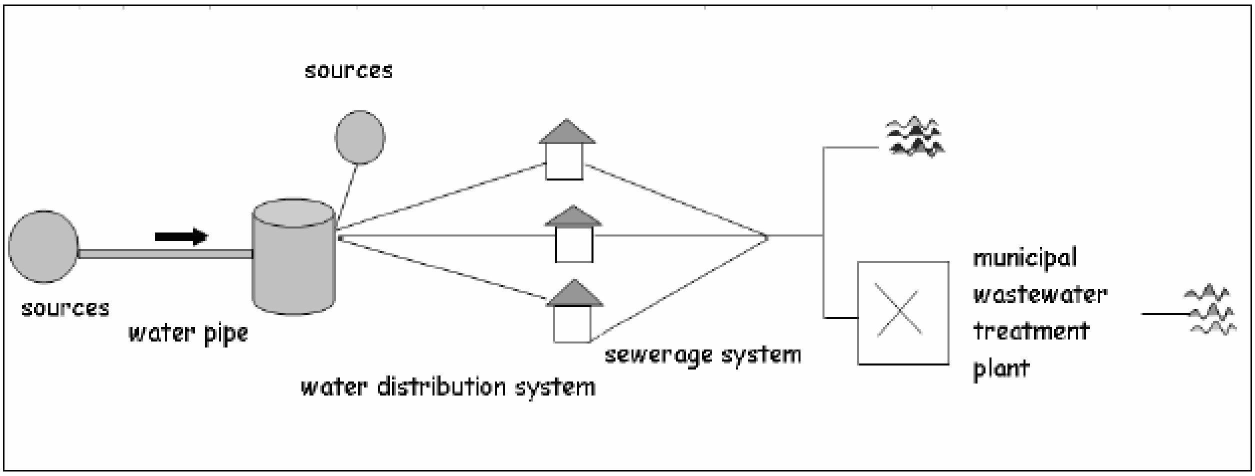
Supply categories are the same as the abstraction ones but, in this case, data refer to public water supplied to the different sectors and to self supply for the other categories.
The wastewater section of JQ-IW includes indicators on population connected to urban wastewater treatment plants (UWWTP), sewage sludge production and disposal, generation and discharge of wastewater. About plants characteristics the JQ collects the number of UWWTP's by type of treatment (primary, secondary, tertiary) and for each type of treatment data on the design capacity, the actual occupation and the effluent in terms of BOD (1000 kg O_2/d) are requested.

The subjects dealt with in the Inland Waters section considered in this paper are water resources related to public supply and wastewater treatment.

3. Istat Water Surveys System contents

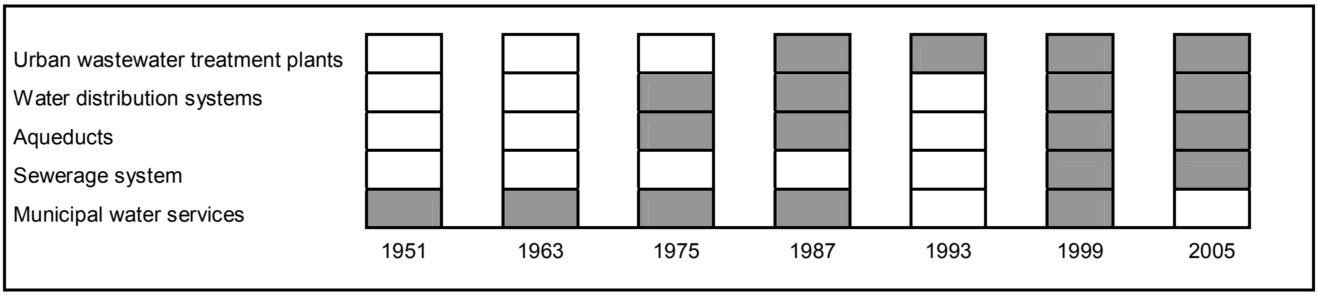
The Istat Water Surveys System (WSS) produces data about municipal water use from abstraction to discharge and about the main characteristics of water services. Data collection requires the definition of a surveys set on: Aqueducts, Public water supply systems, Sewerage systems, Urban wastewater treatment plants, which are the types of facilities furnishing water services. It is also important to ensure the linkage between all the data regarding each of the entities in order to describe the entire phenomena, bearing in mind that water use cycle has to be analysed and observed in an integrated approach (Figure 1).

Figure 1. Public water services scheme



Final respondent units are Water Services Management Companies (WSMC), that are companies in charge of water infrastructures.
From an historical point of view, the Istat water data collections began in 1951 with a survey planned to support a specific sanitation policy; by this time the production process has been developed both for contents and for instruments involved. The time prospectus in Figure 2 shows the collection process, in 1999 the entire surveys system is realized and the name WSS is coined.

Figure 2. Istat water surveys time prospectus



To test important process innovations, aimed at improving data quality and timeliness, a short form has been designed for the last 2005 WSS and only a core data set has been observed. In particular water amount data were observed as total not considering the specific classifications.

In the sub-surveys Aqueducts and Water supply systems, data on public water abstraction for potable use, collected as a whole in 2005, in 1999 WSS census were collected by type of source considering the following items:

- springs
- wells
- water courses
- natural lakes
- artificial reservoirs
- sea and transitional surface water

where ground water comprises springs and wells and surface water is related to the other type of sources; the last source is obviously classified as not fresh water.
Public water supply system sub-survey collects data also on water use. The water use classification adopted in 1999, so to guarantee data provision to Eurostat, was:

- domestic sector
  - households
  - other activities
- production sector
  - industry
  - agriculture
- others uses

In 2005 WSS form, water supplied was observed as a total.
About wastewater, the sub-surveys on Sewerage systems and UWWTP's produce data about the degree of treatment of wastewater collected in sewerage systems (no treatment, partial treatment and complete treatment) and for each UWWTP the type of treatment, the design capacity and the actual occupation in terms of BOD, both in 1999 and 2005 data collection.

4. Testing innovations in the last survey

The 2005 WSS short form, considering a core data set, has been defined to test mainly two important process innovations: the sample design and the use of Computer Assisted Interviewing techniques (CAI), as described in par. 5.
Regarding the first innovation, with the aim of reducing the survey cost, the census approach - traditionally carried out by Istat since 1951 until 1999 - with an about ten-years periodicity, has been replaced by a sampling strategy.
It has to be said that at the moment in Italy a water service reform is in progress with the aim of having an integrated water services management, from abstraction to discharge, for a wide and homogenous portion of territory. Not all the Italian regions have put in practice this reform yet, so, many water utilities provide services in a single municipality and some big water management companies serve multi-jurisdictional areas, which allows to benefit from economies of scale. However privatization of water companies has been recently contested by the so called 'third sector' for poor water quality, increasing prices, and ethical concerns with the result of slowing down the reform process. Then, from the statistical point of view, it is important to monitor the situation and to define flexible data collection procedures.
The survey design strategy adopted for the WSS 2005 edition selected municipalities (the sample frame at the first stage sampling) according to a management criterion: municipalities whose water services were managed by small companies were selected with probability $p<1$, while municipalities grouped and managed by big companies were chosen with probability $p=1$. In this way the managed areas maintain their integrity and data provision is simpler for responding units. At the second stage all WSMC operating in the sampled municipalities were selected.

Following this criterion, the sampling approach in surveying WSMC is tending to move towards the census approach as the water services management is concentrating in few companies according to the service reform.
The advantage of the sampling survey, with respect to census survey, is to reduce cost but the following topics have to be considered:

- because of the high variability of the observed variables (water amounts) the final sampling rate n/N has to be relatively high to produce acceptable sampling errors; in the WSS 2005 the sampling rate for the first stage was 61,3%.
- dimension N of the final sampling frame is not so high (less than 6000 WSMC);
- additional cost for each further unit sampled are not so influent with respect to the initial investment;
- statistical inference process produces data at an aggregated geographical level (regions for the WSS 2005) not useful for more specific environmental analysis. Water resources and water services are distributed in space according to geological factors and human settlement dynamics (as an example see Fig. 3). Even at a small scale there can be a great variability of the phenomenon observed as WSS 1999 data analysis shows (see Ciarallo et al., 2005 and Di Gessa et al., 2006) for spatial statistical analysis using WSS 1999 water abstraction data). In the WSS the minimum geographical scale at which data are observed is municipality. In the past survey editions some questions about geographical coordinates of UWWTP and discharge points were introduced but the quality of data provided and the non response rate suggested the current solution. The municipal geographical level seems to be the best compromise between data relevance and production process efficiency. At this level it is also possible to produce good data estimation at river basins level that is the suitable unit for hydrological analysis.

Operatively the 2005 WSS has been organised in two steps: the first step consisted in a Preliminary Survey (PS) on municipalities and on local water services regulation authorities (*Autorità di Ambito Territoriale Ottimale*), with the aim of updating respondents list. Information resulting from this step also define the data set for the sampling procedure.
In the second step a Definitive Survey (DS) was carried out directly on WSMC's collecting data on water abstraction, water supply, degree of treatment of sewerage system, municipal wastewater treatment plants.

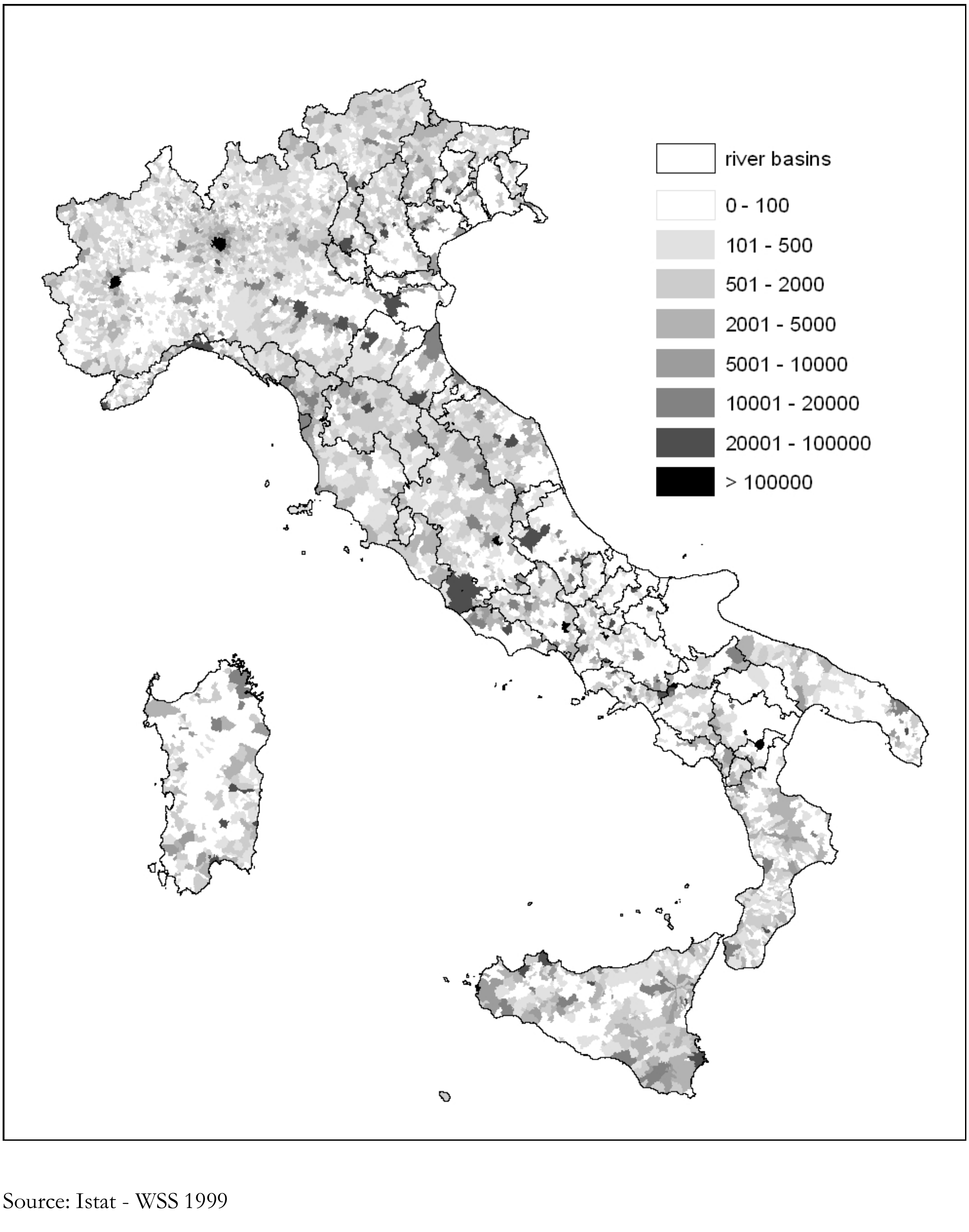
5. The data capturing application design aimed at quality

As already mentioned, another important innovation of 2005 WSS regards the use of CAI techniques.
The previous survey run in 1999 was carried out using self administrated paper questionnaires, while in the WSS 2005 sample survey Computer Assisted Telephone Interviewing (CATI) and Computer Assisted Personal Interviewing (CAPI) were adopted (Couper et al., 1998).
In particular, the Preliminary survey was held with CATI, while the Definitive survey was held integrating the two techniques: CATI and CAPI.
The choice of using interviewing techniques with the presence of the interviewer was taken in order to minimise the total non response rate; in particular, with CATI the call scheduler was set to obtain this result, also encouraging the use of appointments so as to conclude the interview. In addition, the training of interviewers on the meaning of technical concepts used in the questionnaire was done with particular care so to reduce as much as possible the risk of errors due to misunderstanding of definitions.
The decision to use CATI and CAPI was mainly due to the possibilities provided by these techniques of preventing non sampling errors through some specific characteristics they have, i.e.:

a) the data entry is not required, as for paper questionnaires, so the errors proper to this phase are avoided
b) the routing errors are absolutely prevented
c) some checking rules can be managed interactively during the interview

coded during the interview (assisted coding).

Fig. 3. Water abstraction for public water supply in Italy at municipal level (thousands of m^3) Year 1999



Source: Istat - WSS 1999

The decision to use CATI and CAPI was mainly due to the possibilities provided by these techniques of preventing non sampling errors through some specific characteristics they have, i.e.:

e) the data entry is not required, as for paper questionnaires, so the errors proper to this phase are avoided
f) the routing errors are absolutely prevented

g) some checking rules can be managed interactively during the interview
h) categorical data and textual variables referred to an archive coded list can be directly coded during the interview (assisted coding).

The listed potentialities have been exploited in planning and developing the electronic questionnaire (EQ).

In particular the definition of the plan of edits to be included in EQ and their treatment have been done so as to balance the data quality (quality of data to be collected) and the fluency of the interview; all the tools useful to make the interview agreeable for respondents and to reduce respondent burden have been implemented (customisation of questions wording and error messages wording); particular care has been taken of the screen layout, of the error message management and of the helps for the interviewer to make his job as easy as possible.

The care taken in planning and developing the software procedure has been possible thanks to the fact that so called *in-house strategy* has been put in practice for the WSS, which consists in relying on a private company for the call centre, the selection of interviewers and to carry out the interviews, but in giving it all the software procedure, developed in Istat, to manage the data capturing phase (Macchia and Murgia, 2004), concerning:

- the calls scheduler,
- the electronic questionnaire,
- the indicators to monitor the interviewing phase.

The procedure integrates different software packages, but the core is developed with the Blaise system (produced by Statistics Netherlands and already used by a lot of National Statistics Administrations for data capturing carried out with different techniques).

This strategy, already used for other CATI surveys in Istat, guaranteed a strict cooperation between the responsible of the survey and the techniques experts during the EQ planning phase and also a considerable human resource investment for the developing phase, which a private company would have not provided because the limited number of interviews to be completed would have not made up for the costs.

Regarding the prevention of errors, the edit rules implemented in the EQ were treated alternatively in two different ways:

- so as to completely avoid the acceptance of 'wrong' values (*hard* mode: when a 'wrong' value is keyed, an error message is shown and the interviewer asks the respondent to correct his answer);
- so as to minimise the risk of accepting 'wrong' values (*soft* mode: when a 'wrong' value is keyed, an error message is shown and the interviewer asks the respondent to confirm his answer and, if this is the case, the software accepts it).

As already mentioned, the integration of two techniques, CATI and CAPI, was used for the Definitive Survey. As a matter of fact, considering the variability in the amount of data each unit had to provide, respondents have been shared in two groups using scores depending on the number of plants managed. Units having a score under an empirical threshold value, corresponding to a level of sustainability of the telephone interview, were interviewed using CATI technique, for the others the CAPI was used. The same Electronic Questionnaire (EQ) was used for both groups, reducing in this way the effect of the technique. It was given a modular design to the EQ, so that each company was asked only the modules related to types of infrastructure managed, replacing in this way the four paper questionnaires used in the 1999 survey (*Aqueducts, Water distribution systems, Sewerage systems, Municipal wastewater treatment plants*) and guaranteeing data integration. The WSMC was asked to confirm or not the predefined list of plants managed, resulting from the Preliminary survey, so to easily access to the following pertinent questions.

As regards the Preliminary Survey step, Municipalities and Local water services regulation authorities, interviewed with the CATI technique, were asked mainly about qualitative variables such as lists of water pipes (denomination and municipality in which water abstraction is located), water distribution systems and municipality served, sewerage systems and municipality served, municipal wastewater treatment plants (denomination and address). Furthermore for each infrastructure they had to provide the corresponding management company.

The EQ planned for this step includes 30,000 variables and 52 checking rules. This huge number of variables is a theoretical value defined considering the maximum number of different entities which

could be associated to each respondent/form (for instance each respondent can declare from 1 to *n* aqueducts and each aqueduct is associated with a certain number of variables, so any respondent/form reaches this maximum value for all the entities). The checking rules regard, above all, the consistence of the collected information with those of the 1999, in order to avoid wrong linkages of management companies to municipalities and to plants and to point out real changes with respect to the 1999 census, considering also the possible entry of new management companies or new plants.
The second step of the data collection procedure involved mainly quantitative variables about amount of water, asked depending on filter questions aimed at customizing the flow of the questionnaire on the base of the answers to be provided. Water management companies were interviewed using the same EQ submitted both with the CATI and CAPI techniques, as described above.
The modular design of the application allowed respondent to access only the pertinent part of the questionnaire. According to the type of plants managed, the unit was asked to provide data concerning the four modules: *Aqueducts, Water distribution systems, Sewerage systems, Municipal wastewater treatment plants*. For each module a predefined list of plants associated to the water company was proposed on the computer screen and the interviewer, after asking confirmation of the list, carried on with the specific questions related to each plant. For the module *Aqueducts*, data concerned amount of *Water abstracted*, associated with municipality in which the withdrawal source is positioned and *Water treated* volumes. In the *Water distribution system* module data on *Water supplied* to each municipalities served were asked. The module on *Sewerage systems* included a question about the *Degree of treatment* of the wastewater collected in each municipality served and for each wastewater treatment plant listed in the corresponding module, the interviewer had to ask for data about *Type of treatment* (presence/absence of a list of treatment units), *Design capacity* (BOD), *Actual occupation* (BOD).
The EQ included 2849 variables and 52 checking rules. Concerning the checking rules, during these interviews, the variables regarding amounts of water were controlled with respect to their coherence with the values of the 1999 archive, preventing in this way from substantial data capturing errors. In practice, these values were considered as not coherent if they differ from those of 1999 survey over a predefined limit.

6. Results and discussion

In WSS 1999, a frequent cause of error on quantitative variables regarded the coherence between the quantitative value provided by the respondent and the unit of measurement requested (thousands of cubic meters). The most evident type of error was the 1000-factor unit error detected analysing suitable models (per capita water supplied distribution, water invoiced vs water supplied distribution). Interesting probabilistic approaches based on mixture modelling were developed for the localisation of errors (Di Zio et al., 2005).
The use of the EQ in the WSS 2005 largely prevented this kind of error because the interviewer could select the declared unit of measurement used by companies to store their data on water volumes among many available units (m3/y, l/s, l/h,…). In this way, it has been possible to avoid misunderstanding and risks of error in the conversion operation. Comparisons among current values and 1999 archive values within checking rules were performed directly by the software, operating the suitable conversions (archive data were expressed in thousands of cubic meters).
Also concerning textual variables the new data capturing approach deeply improved the production process in terms of timeliness and data accuracy. For the 1999 WSS, pre-coded list of WSMC and pre-coded list of aqueducts were not available and the textual data were collected using the self-administrated paper questionnaires. In the Preliminary survey conducted on municipalities to survey water infrastructures and their management companies, in case a WSMC served more than one municipality, its name was present, usually in not identical ways, in as many questionnaires as the number of municipalities served. Analogously for aqueducts (usually named with local toponyms) when they abstracted water from more municipalities. An ad hoc procedure had to be implemented to identify sets of similar names, equivalence classes, in which a representative text should be selected and assigned as standard coded text.

In the last WSS, using the 1999 coded lists, the assisted coding function provided by the *Blaise* system searches the textual current response in the list of the pre-coded stored descriptions. If an identical description is not found in the pre-coded list, the system performs the match through a texts treatment algorithm based on '*trigrams*' analysis (three consecutive letters of each word), showing in the screen a list of similar descriptions, so leaving the interviewer choosing the standard coded alphanumerical string, which has to be confirmed by the respondent, before being selected.
If any similar description is not found in the pre-coded list, the interviewer has the possibility to insert a new string. In this way the residual checking phase is drastically restricted only to new WSMC and new aqueducts. This coding procedure is not necessary for wastewater treatment plants because their identification is composed not only by the name but also by the address and so duplication errors are simply avoided. For the remaining infrastructure types, sewerage systems and water distribution systems, the identification name corresponds to municipality served so the coding procedure in this case is not affected by errors because it utilizes the well defined municipalities coded list.
Strategies implemented in the WSS 2005 EQ to prevent from non sampling errors are summarised in the scheme in Table 3.

Table 3. Preventions of non sampling errors in the WSS 2005

Main types of error	**Strategies to prevent errors**	**Comments/Results**
Duplication in the WSMC list collected in the Preliminary survey	Assisted coding based on the 1999 WSS database.	The error was completely prevented for all the collected WSMC already included in the pre-coded archive.
Duplication in the aqueducts list collected in the Preliminary survey	Assisted coding based on the 1999 WSS database.	The error was completely prevented for all the collected aqueducts already included in the pre-coded archive.
Linkage among municipalities, WSMC and plants	Queries to a database built according to the known relationships among the entities were made during the interview to control the consistence of responses.	The consistence of the relationships was assured by *hard* controls.
Prevention of partial non responses	When a phenomenon was declared by the respondent, all the related answers had to be given (i.e. in presence of potable treatment, the respondent has to specify the amount of water treated).	All these controls have been managed in *hard* mode: if the respondent could not immediately provide the requested datum, he was encouraged to give an appointment, so as to be able to conclude the interview.
Measurement unit errors	The respondent could select the preferred unit among a pre-coded list. The water amount data were checked interactively during the interview with respect to their coherence with the measurement unit declared and with the values of the 1999 archive (admitting an estimated interval).	All the controls regarding comparison with 1999 archive have been treated in *soft* mode, considering the evolution of the phenomenon.
Consistency with respect to technical constraints	Violations of technical constraints among quantitative variables were controlled during the interview (i.e. the amount of water supplied had to be equal or less than the amount of water feeding the distribution system).	*Hard* controls were implemented to prevent from this type of error.

7. Conclusions

The complex framework of the questionnaires, hidden to respondents, and the amount of controls managed during the interviews did not affect negatively the response rates because they was balanced by the fluency of the interviews and the care taken in the interviewers training and monitoring.

The response rates obtained both for CATI and for CAPI were very high showing the suitability of the adopted approach; the response rate is always upper than 95% as described in Table 4 with small refusal rates.

Table 4. Indicators of the data collection process in WSS 2005

Survey	*Nr. of interviews*	*Interviews' length*	*Days of interview*	*Response rate*	*Refusal rate*
Preliminary Survey CATI	1320	9'03''	16	99.8%	0.1%
Definitive Survey CATI	1423	9'36''	24	96.3%	0.3%
Definitive Survey CAPI	71	-	…	95.8%	4.2%

Regarding to data flow management process, the use of a unique EQ for the Definitive survey brought many advantages both for timeliness and data accuracy with respect to paper questionnaires used for the previous survey.
The good production process timeliness for the Preliminary survey was basic as resulting data defined the personalised lists for the Definitive survey, so the introduction of the CATI technique for the first survey step prevented information from becoming obsolescent.
For the next survey, to be carried out in 2009 for collecting data referred to 2008, the past experience will allow to better define the best data collection plan.

Acknowledgements

The WSS 2005 data collection and the following opportunity of experimenting innovations was possible thanks to the partial support of the Department for Development and Cohesion Policies of the Ministry of Economy and Finance, co-financed Operational Program 2004.
The present paper is financially supported by MIUR COFIN 2006 project Hierarchical Modelling of Spatial And Space-Time Interactions in Environmental Data Series.

References

Amand, M. and R. Montgomery (2001) *Water indicators and data needs: revision of the joint OECD/Eurostat questionnaire on water resources, abstraction and use*, paper presented at *Conference of European Statisticians, Joint ECE/Eurostat Work Session on Methodological Issues of Environment Statistics*, Ottawa, Canada 1-4 October 2001.

Ciarallo, M., A., G. Di Bella, G. Jona Lasinio, M. Remoli, S. Tersigni (2005) *Spatial Analysis of Potable Use Water Resources in Italy: Water Surveys System Results*, *Proceedings of the 2005 Intermediate Congress of the Italian Statistical Society, Statistics and Environment*, Messina, Italy, 21 – 23 September.

Couper, M., P., R. P. Baker, J. Bethlehem, C. Z. F. Clark, W., L. Nicholls II, J. M. O'Reilly (Eds.) (1998) *Computer Assisted Survey Information Collection*, New York: Wiley,.1998.

Di Gessa, G., G. Di Bella, G. Jona Lasinio, D. Ottaviani (2006) *Classification Tree Analysis on the Water Abstraction Making-Decision Process*, paper presented at the *Workshop on Spatial Data Methods for Environmental and Ecological Processes*, Baia delle Zagare, Foggia, Italy, 14-15 September 2006.

Di Zio, M., U. Guarnera, O. Luzi, (2005) *Editing Systematic Unit Measurement Errors through Mixture Modelling*, *Survey Methodology* 31, (2005): 53-63.

Lammer, J. P. M., I. F. E. Kuijpers, J. C. Akkerboom (1997) *Assessment of Quality Criteria in Redesign of Computer Assisted Survey Data Collection*, Bulletin of ISI, 51st Session, book 1, Istanbul.

Macchia, S. and M. Murgia (2004) *Advantages in terms of quality of the in-house strategy for CATI surveys*, paper presented at *Q2004 International Conference on Quality in Official Statistics*, Mainz, May 24-26 2004.

Contributi ISTAT(*)

1/2004 – Marcello D'Orazio, Marco Di Zio e Mauro Scanu – *Statistical Matching and the Likelihood Principle: Uncertainty and Logical Constraints*

2/2004 – Giovanna Brancato – *Metodologie e stime dell'errore di risposta. Una sperimentazione di reintervista telefonica*

3/2004 – Franco Mostacci, Giuseppina Natale e Elisabetta Pugliese – *Gli indici dei prezzi al consumo per sub popolazioni*

4/2004 – Leonello Tronti – *Una proposta di metodo: osservazioni e raccomandazioni sulla definizione e la classificazione di alcune variabili attinenti al mercato del lavoro*

5/2004 – Ugo Guarnera – *Alcuni metodi di imputazione delle mancate risposte parziali per dati quantitativi: il softaware* Quis

6/2004 – Patrizia Giaquinto, Marco Landriscina e Daniela Pagliuca – *La nuova funzione di analisi dei modelli implementata in Genesees v. 3.0*

7/2004 – Roberto Di Giuseppe, Patrizia Giaquinto e Daniela Pagliuca – *MAUSS (Multivariate Allocation of Units in Sampling Surveys): un software generalizzato per risolvere il problema dell' allocazione campionaria nelle indagini Istat*

8/2004 – Ennio Fortunato e Liana Verzicco – *Problemi di rilevazione e integrazione della condizione professionale nelle indagini sociali dell'Istat*

9/2004 – Claudio Pauselli e Claudia Rinaldelli – *La valutazione dell'errore di campionamento delle stime di povertà relativa secondo la tecnica Replicazioni Bilanciate Ripetute*

10/2004 – Eugenio Arcidiacono, Marina Briolini, Paolo Giuberti, Marco Ricci, Giovanni Sacchini e Giorgia Telloli – *Procedimenti giudiziari, reati, indagati e vittime in Emilia-Romagna nel 2002: un'analisi territoriale sulla base dei procedimenti iscritti nel sistema informativo Re.Ge.*

11/2004 – Enrico Grande e Orietta Luzi – *Regression trees in the context of imputation of item non-response: an experimental application on business data*

12/2004 – Luisa Frova e Marilena Pappagallo – *Procedura di now-cast dei dati di mortalità per causa*

13/2004 – Giorgio DellaRocca, Marco Di Zio, Orietta Luzi, Emanuela Scavalli e Giorgia Simeoni – *IDEA (Indices for Data Editing Assessment): sistema per la valutazione degli effetti di procedure di controllo e correzione dei dati e per il calcolo degli indicatori SIDI*

14/2004 – Monica Pace, Silvia Bruzzone, Luisa Frova e Marilena Pappagallo – *Review of the existing information about death certification practices, certificate structures and training tools for certification of causes of death in Europe*

15/2004 – Elisa Berntsen – *Modello Unico di Dichiarazione ambientale: una fonte amministrativa per l'Archivio delle Unità Locali di Asia*

16/2004 – Salvatore F. Allegra e Alessandro La Rocca – *Sintetizzare misure elementari: una sperimentazione di alcuni criteri per la definizione di un indice composto*

17/2004 – Francesca R. Pogelli – *Un'applicazione del modello "Country Product Dummy" per un'analisi territoriale dei prezzi*

18/2004 – Antonia Manzari – *Valutazione comparativa di alcuni metodi di imputazione singola delle mancate risposte parziali per dati quantitativi*

19/2004 – Claudio Pauselli – *Intensità di povertà relativa: stima dell'errore di campionamento e sua valutazione temporale*

20/2004 – Maria Dimitri, Ersilia Di Pietro, Alessandra Nuccitelli e Evelina Paluzzi – *Sperimentazione di una metodologia per il controllo della qualità di dati anagrafici*

21/2004 – Tiziana Pichiorri, Anna M. Sgamba e Valerio Papale – *Un modello di ottimizzazione per l'imputazione delle mancate risposte statistiche nell'indagine sui trasporti marittimi dell'Istat*

22/2004 – Diego Bellisai, Piero D. Falorsi, Annalisa Lucarelli, Maria A. Pennucci e Leonello G. Tronti – *Indagine pilota sulle retribuzioni di fatto nel pubblico impiego*

23/2004 – Lidia Brondi – *La riorganizzazione del sistema idrico: quadro normativo, delimitazione degli ambiti territoriali ottimali e analisi statistica delle loro caratteristiche strutturali*

24/2004 – Roberto Gismondi e Laura De Sandro – *Provisional Estimation of the Italian Monthly Retail Trade Index*

25/2004 – Annamaria Urbano, Claudia Brunini e Alessandra Chessa – *I minori in stato di abbandono: analisi del fenomeno e studio di una nuova prospettiva d'indagine*

26/2004 – Paola Anzini e Anna Ciammola – *La destagionalizzazione degli indici della produzione industriale: un confronto tra approccio diretto e indiretto*

27/2004 – Alessandro La Rocca – *Analisi della struttura settoriale dell'occupazione regionale: 8° Censimento dell'industria e dei servizi 2001 7° Censimento dell'industria e dei servizi 1991*

28/2004 – Vincenzo Spinelli e Massimiliano Tancioni – *I Trattamenti Monetari non Pensionistici: approccio computazionale e risultati della sperimentazione sugli archivi INPS-DM10*

29/2004 – Paolo Consolini – *L'indagine sperimentale sull'archivio fiscale modd.770 anno 1999: analisi della qualità del dato e stime campionarie*

1/2005 – Fabrizio M. Arosio – *La stampa periodica e l'informazione on-line: risultati dell'indagine pilota sui quotidiani on-line*

2/2005 – Marco Di Zio, Ugo Guarnera e Orietta Luzi – *Improving the effectiveness of a probabilistic editing strategy for business data*

3/2005 – Diego Moretti e Claudia Rinaldelli – *EU-SILC complex indicators: the implementation of variance estimation*

4/2005 – Fabio Bacchini, Roberto Iannaccone e Edoardo Otranto – *L'imputazione delle mancate risposte in presenza di dati longitudinali: un'applicazione ai permessi di costruzione*

5/2005 – Marco Broccoli – *Analisi della criminalità a livello comunale: metodologie innovative*

6/2005 – Claudia De Vitiis, Loredana Di Consiglio e Stefano Falorsi – *Studio del disegno campionario per la nuova rilevazione continua sulle Forze di Lavoro*

7/2005 – Edoardo Otranto e Roberto Iannaccone – *Continuous Time Models to Extract a Signal in Presence of Irregular Surveys*

(*) ultimi cinque anni

8/2005 – Cosima Mero e Adriano Pareto – *Analisi e sintesi degli indicatori di qualità dell'attività di rilevazione nelle indagini campionarie sulle famiglie*

9/2005 – Filippo Oropallo – *Enterprise microsimulation models and data challenges*

10/2005 – Marcello D' Orazio, Marco Di Zio e Mauro Scanu – *A comparison among different estimators of regression parameters on statistically matched files through an extensive simulation study*

11/2005 – Stefania Macchia, Manuela Murgia, Loredana Mazza, Giorgia Simeoni, Francesca Di Patrizio, Valentino Parisi, Roberto Petrillo e Paola Ungaro – *Una soluzione per la rilevazione e codifica della Professione nelle indagini CATI*

12/2005 – Piero D. Falorsi, Monica Scannapieco, Antonia Boggia e Antonio Pavone – *Principi Guida per il Miglioramento della Qualità dei Dati Toponomastici nella Pubblica Amministrazione*

13/2005 – Ciro Baldi, Francesca Ceccato, Silvia Pacini e Donatella Tuzi – *La stima anticipata OROS sull'occupazione. Errori, problemi della metodologia attuale e proposte di miglioramento*

14/2005 – Stefano De Francisci, Giuseppe Sindoni e Leonardo Tininini – *Da Winci/MD: un sistema per data warehouse statistici sul Web*

15/2005 – Gerardo Gallo e Evelina Palazzi – *I cittadini italiani naturalizzati: l'analisi dei dati censuari del 2001, con un confronto tra immigrati di prima e seconda generazione*

16/2005 – Saverio Gazzelloni, Mario Albisinni, Lorenzo Bagatta, Claudio Ceccarelli, Luciana Quattrociocchi, Rita Ranaldi e Antonio Toma – *La nuova rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*

17/2005 – Maria Carla Congia – *Il lavoro degli extracomunitari nelle imprese italiane e la regolarizzazione del 2002. Prime evidenze empiriche dai dati INPS*

18/2005 – Giovanni Bottazzi, Patrizia Cella, Giuseppe Garofalo, Paolo Misso, Mariano Porcu e Marianna Tosi – *Indagine pilota sulla nuova imprenditorialità nella Regione Sardegna. Relazione Conclusiva*

19/2005 – Fabrizio Martire e Donatella Zindato – *Le famiglie straniere: analisi dei dati censuari del 2001 sui cittadini stranieri residenti*

20/2005 – Ennio Fortunato – *Il Sistema di Indicatori Territoriali: percorso di progetto, prospettive di sviluppo e integrazione con i processi di produzione statistica*

21/2005 – Antonella Baldassarini e Danilo Birardi – *I conti economici trimestrali: un approccio alla stima dell'input di lavoro*

22/2005 – Francesco Rizzo, Dario Camol e Laura Vignola – *Uso di XML e WEB Services per l'integrazione di sistemi informativi statistici attraverso lo standard SDMX*

1/2006 – Ennio Fortunato – *L'analisi integrata delle esigenze informative dell'utenza Istat: Il contributo del Sistema di Indicatori Territoriali*

2/2006 – Francesco Altarocca – *I design pattern nella progettazione di software per il supporto alla statistica ufficiale*

3/2006 – Roberta Palmieri – *Le migranti straniere: una lettura di genere dei dati dell'osservatorio interistituzionale sull'immigrazione in provincia di Macerata*

4/2006 – Raffaella Amato, Silvia Bruzzone, Valentina Delmonte e Lidia Fagiolo – *Le statistiche sociali dell'ISTAT e il fenomeno degli incidenti stradali: un'esperienza di record linkage*

5/2006 – Alessandro La Rocca – *Fuzzy clustering: la logica, i metodi*

6/2006 – Raffaella Cascioli – *Integrazione dei dati micro dalla Rilevazione delle Forze di Lavoro e dagli archivi amministrativi INPS: risultati di una sperimentazione sui dati campione di 4 province*

7/2006 – Gianluca Brogi, Salvatore Cusimano, Giuseppina del Vicario, Giuseppe Garofalo e Orietta Patacchia – *La realizzazione di Asia Agricoltura tramite l'utilizzo di dati amministrativi: il contenuto delle fonti e i risultati del processo di integrazione*

8/2006 – Simonetta Cozzi – *La distribuzione commerciale in Italia: caratteristiche strutturali e tendenze evolutive*

9/2006 – Giovanni Seri – *A graphical framework to evaluate risk assessment and information loss at individual level*

10/2006 – Diego Bellisai, Annalisa Lucarelli, Maria Anna Pennucci e Fabio Rapiti – *Feasibility studies for the coverage of public institutions in sections N and O*

11/2006 – Diego Bellisai, Annalisa Lucarelli, Maria Anna Pennucci e Fabio Rapiti – *Quarterly labour cost index in public education*

12/2006 – Silvia Montagna, Patrizia Collesi, Florinda Damiani, Danila Fulgenzio, Maria Francesca Loporcaro e Giorgia Simeoni – *Nuove esperienze di rilevazione della Customer Satisfaction*

13/2006 – Lucia Coppola e Giovanni Seri – *Confidentiality aspects of household panel surveys: the case study of Italian sample from EU-SILC*

14/2006 – Lidia Brondi – *L'utilizzazione delle surveys per la stima del valore monetario del danno ambientale: il metodo della valutazione contingente*

15/2006 – Carlo Boselli – *Le piccole imprese leggere esportatrici e non esportatrici: differenze di struttura e di comportamento*

16/2006 – Carlo De Greogorio – *Il nuovo impianto della rilevazione centralizzata del prezzo dei medicinali con obbligo di prescrizione*

1/2007 – Paolo Roberti, Maria Grazia Calza, Filippo Oropallo e Stefania Rossetti – *Knowledge Databases to Support Policy Impact Analysis: the EuroKy-PIA Project*

2/2007 – Ciro Baldi, Diego Bellisai, Stefania Fivizzani, e Marina Sorrentino – *Production of job vacancy statistics: coverage*

3/2007 – Carlo Lucarelli e Giampiero Ricci – *Working times and working schedules: the framework emerging from the new Italian lfs in a gender perspective*

4/2007 – Monica Scannapieco, Diego Zardetto e Giulio Barcaroli – *La Calibrazione dei Dati con R: una Sperimentazione sull'Indagine Forze di Lavoro ed un Confronto con GENESEES/SAS*

5/2007 – Giulio Barcaroli e Tiziana Pellicciotti – *Strumenti per la documentazione e diffusione dei microdati d'indagine: il Microdata Management Toolkit*

6/2007 – AA.VV. – *Seminario sulla qualità: l'esperienza dei referenti del sistema informativo SIDI - 1ª giornata*

7/2007 – Raffaella Cianchetta, Carlo De Gregorio, Giovanni Seri e Giulio Barcaroli – *Rilevazione sulle Pubblicazioni Scientifiche Istat*

8/2007 – Emilia Arcaleni, e Barbara Baldazzi – *Vivere non insieme: approcci conoscitivi al Living Apart Together*

9/2007 – Corrado Peperoni e Francesca Tuzi – *Trattamenti monetari non pensionistici metodologia sperimentale per la stima degli assegni al nucleo familiare*

10/2007 – AA.VV – *Seminario sulla qualità: l'esperienza dei referenti del sistema informativo SIDI - 2ª giornata*

11/2007 – Leonello Tronti – *Il prototipo (numero 0) dell'Annuario di statistiche del Mercato del Lavoro (AML)*

12/2007 – Daniele Frongia, Raffaello Martinelli, Fernanda Panizon, Bruno Querini e Andrea Stanco – *Il nuovo Sistema informatico Altri Servizi. Progetto di reingegnerizzazione dei processi produttivi delle indagini trimestrali di fatturato degli altri servizi*

1/2008 – Carlo De Gregorio, Stefania Fatello, Rosanna Lo Conte, Stefano Mosca, Francesca Rossetti – *Sampling design and treatment of products in Istat centralised CPI surveys*

2/2008 – Mario Abissini, Elisa Marzilli e Federica Pintaldi – *Test cognitivo e utilizzo del questionario tradotto: sperimentazioni dell'indagine sulle forze di lavoro*

3/2008 – Franco Mostacci – *Gli aggiustamenti di qualità negli indici dei prezzi al consumo in Italia: metodi, casi di studio e indicatori impliciti*

4/2008 – Carlo Vaccari e Daniele Frongia – *Introduzione al Web 2.0 per la Statistica*

5/2008 – Antonio Cortese – *La conta degli stranieri: una bella sfida per il censimento demografico del 2011*

6/2008 – Carlo De Gregorio, Carmina Munzi e Paola Zavagnini – *Problemi di stima, effetti stagionali e politiche di prezzo in alcuni servizi di alloggio complementari: alcune evidenze dalle rilevazioni centralizzate dei prezzi al consumo*

7/2008 – AA.VV. – *Seminario: metodi per il controllo e la correzione dei dati nelle indagini sulle imprese: alcune esperienze nel settore delle statistiche strutturali*

8/2008 – Monica Montella – *La nuova matrice dei margini di trasporto*

9/2008 – Antonia Boggia, Marco Fortini, Matteo Mazziotta, Alessandro Pallara, Antonio Pavone, Federico Polidoro, Rosabel Ricci, Anna Maria Sgamba e Angela Seeber – *L'indagine conoscitiva della rete di rilevazione dei prezzi al consumo*

10/2008 – Marco Ballin e Giulio Barcaroli – *Optimal stratification of sampling frames in a multivariate and multidomain sample design*

11/2008 – Grazia Di Bella e Stefania Macchia – *Experimenting Data Capturing Techniques for Water Statistics*