

CENSIMENTI 2010-2011

Resoconto dell'audizione con Direttori e Dirigenti dell'Istat

Roma, 27 febbraio 2009

PARTECIPANTI

ISTAT:

dott. A. Mancini (Responsabile della Direzione centrale dei censimenti generali), dott. G. Carbonetti, dott. F. Crescenzi, dott. F. Lorenzini, dott. M. Fortini, dott.ssa A. Reale, dott. G. Stassi.

Commissione per la Garanzia dell'Informazione Statistica (CoGIS):

prof. A. Chiappetti (Presidente), prof.ssa M. Civardi, prof. E. Del Colle, prof. C. Filippucci
dott.ssa P. Paduano, prof.ssa S. Salvini.

Segreteria tecnico-statistica della CoGIS:

dott.ssa L. Cataldi, dott.ssa C. Panattoni.

Il prof. Chiappetti introduce l'incontro.

Il prof. Filippucci esprime preoccupazioni sul possibile mancato finanziamento da parte del governo delle operazioni censuarie e chiede chiarimenti in merito.

Il dott. Mancini ricorda che i prossimi Censimenti della popolazione e dell'agricoltura sono soggetti a Regolamento comunitario che fissa, tra l'altro, il periodo di riferimento e le scadenze per la consegna dei dati ad Eurostat. Per rispettare i tempi stabiliti occorrerebbe che il governo emanasse i Regolamenti esecutivi almeno dodici mesi prima dello svolgimento delle operazioni censuarie, in modo tale da permettere alle amministrazioni pubbliche incaricate della rilevazione sul campo di pianificare le proprie operazioni e predisporre gli stanziamenti di bilancio. Ogni ritardo comporta una dilazione dei tempi ed in caso di inadempienza dei regolamenti comunitari, sono previste sanzioni che potrebbero risultare anche molto onerose per il nostro Paese.

Il dott. Mancini afferma inoltre che, mentre il Censimento dell'agricoltura non è ulteriormente restringibile da un punto di vista finanziario, il Censimento della popolazione invece può subire eventuali variazioni di costo in funzione delle diverse scelte strategiche che ruotano su tre punti chiave: (i) diminuzione del fastidio statistico sui rispondenti (ii) aumento della tempestività (iii) semplificazione del lavoro dei Comuni riducendo il numero dei rilevatori, ovviando ai problemi incontrati nella scorsa tornata censuaria in merito alla formazione e al controllo dei rilevatori. A tale proposito, il Prof. Filippucci sottolinea

l'impatto negativo che potrebbe avere la riduzione dei finanziamenti sul lavoro dei rilevatori ricordando che, nel comune di Milano nel 2001, abbandonarono l'incarico, dopo essere stati formati, per mancanza di fondi.

La prof.ssa Civardi ribadisce l'importanza del controllo sul territorio e i Proff. Filippucci e Salvini pongono l'accento sulle modalità di restituzione dei questionari. In merito il dott. Mancini illustra tre aspetti innovativi previsti per il prossimo Censimento della popolazione: (i) le liste comunali anagrafiche (Lac) saranno utilizzate per il mail-out e saranno previsti più canali per la restituzione dei questionari (posta, web, centro di raccolta comunale). Il ritorno e il controllo di tutti i questionari sarà garantito attraverso l'invio dei rilevatori direttamente ai numeri civici per recuperare le mancate risposte, usufruendo dei risultati della Rnc. (ii) La sottocopertura delle anagrafi potrà essere affrontata secondo due strategie alternative: (a) attraverso il recupero sistematico tramite rilevatori. Successivamente alla realizzazione del Censimento le stesse informazioni potrebbero essere utilizzate dai Comuni per l'aggiornamento delle anagrafi; (b) attraverso una rilevazione integrativa su alcune sezioni di censimento, così come avviene per il Censimento tradizionale non assistito da lista. In questo caso – con l'applicazione di un modello cattura/ricattura – si stimerebbero i tassi di sottocopertura che verrebbe recuperata senza intervento del Comune e uniformemente per tutto il territorio nazionale. Il ricorso a tale strategia consentirebbe di effettuare anche un controllo di sopracopertura delle anagrafi comunali. D'altra parte occorre sottolineare che la seconda strategia ha anche un'indubbia conseguenza negativa: una rilevazione del genere non consente, infatti, di individuare i nominativi di coloro che sono sfuggiti al censimento e, di conseguenza, nella fase di confronto tra censimento e anagrafe, la precisa individuazione dei soggetti non rilevati graverebbe interamente sul Comune. (iii) Verranno utilizzate forme differenziate di questionario: *short (SF)* o *medium form (MF)* e *long form (LF)*. I questionari verranno "testati" attraverso l'indagine pilota prevista nel 2009 e si valuterà, in quella occasione, l'impatto delle diverse lunghezze sul tasso di restituzione automatico. La prof.ssa Salvini sottolinea come questa strategia non sia priva di conseguenze, soprattutto di tipo conoscitivo, prima fra tutte, l'impossibilità di disporre di dati sul livello di istruzione e condizione professionale dei cittadini a scala comunale e sub-comunale per tutte quelle aree, appunto, che non entreranno a far parte del campione a cui sarà somministrato il questionario *long-form*. Il censimento è infatti l'unica fonte che consente la ricostruzione esaustiva della classificazione per istruzione e tipologia professionale della popolazione a livello territoriale disaggregato, poiché queste informazioni aggiornate non sono disponibili dall'anagrafe dei comuni.

Il Prof. Filippucci evidenzia alcuni aspetti di criticità che si proporranno, a suo giudizio, durante l'esecuzione del prossimo Censimento generale della popolazione. In particolare:

- l'adozione di più modalità di restituzione dei questionari (raccolta multicanale) necessita di un'attenta predisposizione, da parte dell'Istat, di strategie di revisione da mettere in campo a seguito dell'utilizzo dei diversi canali a disposizione. Tali strategie devono, inoltre, essere adeguatamente testate;
- la strategia generale scelta dall'Istat, ossia quella di utilizzare l'anagrafe comunale (Lac) quale archivio principale, presenta importanti criticità considerando i non pochi errori dai quali risultano affetti i registri anagrafici sui quali a tutt'oggi non sono generalmente stati effettuati appropriati controlli diretti, come in precedenza suggerito, magari attraverso opportune indagini areali, peraltro molto dispendiose anche tenendo conto che il lavoro di analisi¹ che l'Istituto ha svolto in tale direzione appare ad oggi alquanto datato. Un problema analogo si pone per il Censimento dell'Industria e dei servizi e per il relativo utilizzo dell'archivio Asia che in genere è in ritardo di circa un anno e mezzo, quindi con la possibilità che non colga la lista effettiva delle imprese attive nel 2011 a causa sia della crescita della mortalità che si potrebbe manifestare nel corso del 2009-2010 per la crisi economica in atto, sia della nascita di imprese che emergano da tale crisi;
- il sistema di registrazione dovrà essere diverso a seconda del canale di restituzione scelto. In particolare, si richiede se sono state prese in considerazione e adeguatamente testate le due modalità di registrazione attraverso lettura ottica e/o manuale dei questionari che saranno restituiti in forma cartacea e che potrebbero rappresentare una parte cospicua dei questionari compilati. Poiché emerge che non si è presa in considerazione la possibilità di ricorrere alla lettura ottica a causa dei molti errori che risultarono al censimento precedente, il prof. Filippucci sottolinea che anche la registrazione manuale non è affatto esente da errori, al contrario, in passato si optò per la lettura ottica proprio a causa dei molti errori di registrazione. Inoltre, la registrazione manuale comporta tempi molto lunghi, il rischio di significative difformità di performance delle diverse aziende chiamate ad effettuare la registrazione e il rischio di smarrimento dei questionari nella fase di trasmissione a tali aziende. Il prof. Filippucci sottolinea l'importanza di approfondire maggiormente le problematiche relative alla registrazione.

¹ Cfr. *Misure di sottocopertura anagrafica in base alla revisione post-censuaria del 2001* (M.Fortini, G.Gallo)

- il sistema di monitoraggio e controllo delle operazioni del 15° Censimento generale della Popolazione si presenta più complesso rispetto a quello della scorsa tornata censuaria per la coesistenza di più procedure (legate alla multicanalità) e per gli stretti tempi a disposizione e pertanto richiederà una sperimentazione adeguata;
- la strategia generale messa in campo, che prevede l'utilizzo di due diversi questionari (*short-form* SF e *long-form* LF), porta alla determinazione di due diversi errori: campionario e non. Ad oggi sembra che non siano stati adeguatamente prefissati degli obiettivi precisi in merito all'errore campionario che si è disposti ad accettare e rispetto alla definizione di soglie del coefficiente di variazione (CV), al di là delle quali si renda necessario introdurre adeguate strategie di correzione. A ciò si aggiunge che i diversi canali di restituzione dei questionari previsti implicheranno tassi di risposta diversi e diversi tassi di errore. Un aspetto non meno importante, riguarda le difficoltà che un utilizzatore non esperto potrà incontrare nell'interpretare ed usare i dati prodotti in quanto nelle tabelle presentate verranno incrociate variabili affette da errore campionario (le variabili socio economiche, come l'istruzione e la professione, rilevate soltanto tramite questionario LF) con variabili affette solo da errore non campionario (tutte quelle di tipo anagrafico rilevate tramite questionario SF). Per ovviare a queste problematiche l'Istat dovrà formulare anche un'adeguata strategia di comunicazione dei metadati.

La Prof.ssa Civardi e il Prof. Del Colle si associano nell'esprimere preoccupazioni riguardo all'affidabilità dei dati risultanti congiuntamente da una rilevazione censuaria e da una campionaria. Il dott. Fortini ribadisce che l'errore non campionario è molto alto nei censimenti tradizionali mentre è ovviamente più contenuto nelle rilevazioni campionarie: dunque, non deve destare maggiore preoccupazione la presenza di quello campionario che è controllabile attraverso adeguate strategie di campionamento.

La Prof. ssa Civardi mostra, inoltre, le sue perplessità circa l'utilizzo della Lac.

Il dott. Mancini ricorda che esiste un sistema di screening per valutare la tenuta delle anagrafi. Per il prossimo censimento si è optato per una strategia mista che, inizialmente, si avvale della Lac per la distribuzione dei questionari, ma, successivamente, vede impegnati gli UCC nel recupero delle unità sfuggite che potranno avvalersi sia dell'aiuto fornito dal sistema di monitoraggio che delle informazioni raccolte mediante la Rilevazione di numeri civici (Rnc) da effettuare in anticipo rispetto al censimento. Inoltre proprio la Rnc costituirà uno dei "ritorni" più importanti per la nazione quale strumento per l'aggiornamento delle basi

territoriali dei diversi comuni e la geocodificazione delle informazioni anche di fonte amministrativa.

Quanto alle strategie di comunicazione, il dott. Mancini puntualizza che, anche in ottemperanza al principio della trasparenza, la documentazione sulla qualità dei dati verrà anticipata e non proposta solo in sede di sperimentazione, mentre, in relazione al sistema di monitoraggio, ribadisce che si tratta di un sistema ampiamente flessibile che permetterà un collegamento interattivo con i vari UCC. Inoltre permetterà ai Comuni di essere autonomi poiché consentirà di realizzare tabelle di controllo ed altri strumenti di verifica da poter utilizzare all'occorrenza.

In merito alle strategie di campionamento, il dott. Carbonetti riporta brevemente i risultati della sperimentazione² (sui dati del censimento della popolazione del 2001) svolta su 40 Comuni (10% delle famiglie e degli individui censiti al 2001), finalizzata sia a fornire indicazioni circa la definizione della strategia campionaria da adottare per il prossimo censimento, sia a valutare l'impatto delle innovazioni metodologiche proposte sulla produzione delle informazioni censuarie e sul livello di qualità delle tavole di diffusione. Dunque riferisce che, mentre solitamente si parte dal livello di accuratezza prefissato per le stime e, in base a quello, si determina la frazione di campionamento, per questa tornata censuaria non si è potuto operare in questo modo. In una prima fase si è valutato il disegno campionario che meglio può adattarsi al censimento (scegliendo il disegno di campionamento casuale semplice di famiglie da lista). In una seconda fase la sperimentazione si è concentrata sulle tavole di diffusione (contenenti sia le informazioni da produrre a livello comunale e da trasmettere per regolamento europeo, sia quelle riferite alle aree di censimento sub comunali). L'obiettivo è l'accuratezza delle stime (prefissata per ogni cella) inerenti le frequenze relative e assolute sia per le modalità delle variabili LF singole sia per quelle di incrocio con le modalità delle variabili esaustive SF.

L'utente finale avrà a disposizione, per ognuna delle tabelle di dati prodotta, la relativa tabella contenente i coefficienti di variazione (CV), ossia gli errori percentuali medi di ciascuna cella. Rimane ancora aperta la questione su quale sia l'errore percentuale atteso massimo accettabile per poter ritenere soddisfacente il livello di qualità complessiva delle singole tavole statistiche.

Passando all'esame delle bozze dei questionari predisposti per il Censimento della popolazione, la prof.ssa Salvini auspica innanzitutto una scelta che sia a favore della *medium*

² Cfr. *Problemi di accuratezza delle stime da campioni di famiglie in un contesto censuario* (F. Borrelli, G. Carbonetti, L. De Felici)

form piuttosto che della *short form*, in particolare riferendosi a quelle carenze – in precedenza delineate – nel caso in cui si dovessero rilevare le informazioni sul livello di istruzione soltanto sui comuni campionati per la *long-form*. Poi pone l'attenzione sulla struttura dei questionari stessi. Rileva che essi sono poco *user friendly* e, soprattutto, privi di filtri espliciti. Il ricorso a continui rimandi comporta sicuramente dei rischi di inaccuratezza delle risposte da parte dei cittadini che, in gran parte, si troveranno a rispondere ai quesiti censuari privi della presenza di intervistatori. Piuttosto auspica che si possa adottare una struttura simile a quella dei questionari dell'indagine multiscopo, dove i filtri aiutano la compilazione dei questionari stessi. Inoltre, la lettera del Presidente dell'Istat dovrebbe essere a parte, non inserita nel frontespizio del questionario.

Portando come esempio le pagine 2 e 3, fa notare la difficoltà che si incontra nel leggere caratteri troppo piccoli e nel riempire caselle anguste. Per quanto attiene alle problematiche relative all'anonimato e alla riservatezza delle informazioni rilasciate, la prof.ssa Salvini chiede se l'Istat prevede l'utilizzo del lembo staccabile. Affronta poi i problemi legati al *wording*. Data anche la multicanalità delle risposte, occorrerebbe utilizzare un linguaggio che sia comprensibile a tutti, più vicino all'uso comune, senza il continuo rinvio al foglio delle istruzioni. L'eccessivo rimando alle istruzioni, infatti, può generare un grave appesantimento nella compilazione del questionario. Si potrebbe ovviare portando degli esempi esplicativi accanto ai quesiti. Infine, la prof.ssa Salvini chiede se le definizioni adottate nel questionario sono armonizzate con quelle delle altre indagini condotte dall'Istat (es: quella di occupato armonizzata con la RCFL).

Risponde il dott. Stassi che rassicura sul fatto che i questionari all'esame della CoGIS sono solo delle bozze suscettibili di correzioni e miglioramenti, alcuni dei quali già messi in cantiere come l'eliminazione della lettera del Presidente dell'Istat dal frontespizio del questionario. Inoltre, la grafica utilizzata per le bozze dei questionari è quella del 2001 e bisognerà ancora lavorare per produrre il questionario definitivo. Per quanto riguarda le definizioni utilizzate sono in linea con quelle delle altre indagini Istat. Infine, il dott. Stassi conclude affermando che la multicanalità delle risposte pone problemi relativi alla valutazione ex-ante della preparazione di coloro che sceglieranno di rispondere on-line: la realtà è che esistono situazioni variegata e intermedie tra la tipologia di "utente Internet specializzato" e la tipologia di "utente Internet che ha le stesse competenze di chi compila il questionario cartaceo". E di questo l'Istat sta tenendo conto.

Il prof. Filippucci replica che, se il questionario on-line non viene presentato correttamente e secondo la giusta forma, può dar luogo ad abbandoni. E proprio su quest'ultimo argomento chiede se sono state previste modalità di gestione degli abbandoni on-line, ovvero la possibilità per l'utente di abbandonare la compilazione in qualsiasi o determinato momento per optare per le altre forme di restituzione. Il dott. Stassi risponde che l'Istat non ha al momento preso decisioni, ma che tutte le scelte in merito saranno comunicate e illustrate agli utenti nella prima pagina del questionario.

Passando al Censimento dell'agricoltura, il Prof. Filippucci esprime alcune perplessità sorte dalla lettura della documentazione inviata dall'Istat. E, in particolare, prendendo in considerazione i primi risultati desunti dal sistema di monitoraggio della rilevazione campionaria di controllo della copertura e della qualità del prototipo di registro statistico delle aziende agricole, emerge che il 31,71% delle unità rilevate in lista sono risultate svolgere attività agricola³. Non appare, invece, ben chiaro come, terminata la rilevazione, si utilizzerà il dato campionario per predisporre la lista precensuaria.

Il dott. Lorenzini, illustrando brevemente il disegno di campionamento della rilevazione, puntualizza che la lista prototipale del registro statistico delle aziende agricole è esente dal rischio di sottocopertura in quanto risulta sovradimensionata a causa dell'inclusione dei soggetti registrati negli archivi "Redditi agrari" (di fonte Agenzia delle Entrate) e "Catasto dei terreni ad uso agricolo" (di fonte Agenzia del Territorio). La rilevazione, dunque, si propone essenzialmente di eliminare l'errore di sovracopertura che potrebbe invece risultare consistente per la presenza di unità non appartenenti al campo di osservazione statistica o di unità non più attive. Ad oggi la lista prototipale è costituita da circa 8 milioni di unità; essa, alla fine della rilevazione, dovrebbe restituire una lista precensuaria, contenente circa 2 milioni di unità (aziende agricole). La rilevazione prevede l'intervista diretta delle unità in lista ed è organizzata in due stadi di campionamento. Al primo stadio sono stati selezionati 80 comuni, di concerto tra l'Istat e le Regioni, in modo da rappresentare le diverse specializzazioni agricole provinciali; al secondo stadio vengono costituite, sugli 80 comuni, 5 sottoliste (mutuando il metodo utilizzato per la costituzione dell'archivio ASIA Imprese) per un campione complessivo di circa 16.000 unità la cui esatta natura viene verificata sul campo.

Il dott. Lorenzini sottolinea che anche per il censimento dell'agricoltura si prevede l'uso della tecnica di rilevazione multicanale, per cui saranno predisposte diverse strategie di controllo e correzione dei questionari; ad esempio, per i questionari compilati on-line, si

³ I dati della rilevazione sono aggiornati al 31/12/2008 e rilevano una percentuale complessiva di restituzione a livello nazionale pari al 10,05%.

prevede anche l'eventuale ritorno del rilevatore nel caso in cui si dovessero evidenziare dei valori anomali rispetto a determinati valori di controllo che l'Istat dovrà stabilire. Se il canale scelto sarà quello della restituzione del cartaceo, il questionario verrà inviato ad un'apposita ditta per la registrazione, così come è avvenuto per i passati censimenti.

In generale, quanto ai tassi di utilizzo dei diversi canali di restituzione, si è stimato, sulla base di studi comparativi con altri paesi e con altre indagini, che circa un 30% degli intervistati utilizzerà il canale postale, circa un 40% il centro di raccolta comunale o il rilevatore, mentre il canale web riguarderà circa un 20% (considerando che circa il 40% delle famiglie italiane è risultata in possesso di un pc). Il Prof. Filippucci ribadisce la sua preoccupazione, dato l'alto tasso stimato (circa il 70%) di questionari cartacei, rispetto alla scelta (almeno momentanea) dell'Istat di rinunciare alla lettura ottica (OCR) a favore della registrazione manuale. Infatti, i tempi di registrazione potranno essere molto lunghi provocando un notevole ritardo sulla restituzione dell'informazione finale. La dott.ssa Reale ricorda che tale scelta è scaturita dalle problematiche che, nei trascorsi censimenti, la lettura ottica ha causato, sia per il censimento dell'agricoltura (difficoltà nell'individuazione delle aree), sia per molte delle variabili utilizzate nel censimento della popolazione. Sulla questione, comunque, l'Istat non esclude la possibilità di utilizzare tecnologie OCR per la lettura ottica. Questa soluzione è al vaglio del comitato informatica per i censimenti.

Il prof. Filippucci rileva, inoltre, che sarebbe utile che l'Istat nella progettazione del questionario considerasse la possibilità di ricorrere a strategie di progettazione cognitiva, come del resto ormai viene fatto in molte importanti agenzie statistiche nazionali, sia per ottenere la migliore comprensibilità possibile delle domande da parte della popolazione, sia per definire un'appropriata struttura del questionario.

In conclusione della riunione, i Commissari della CoGIS fanno presente che restano ancora aperte alcune questioni riguardanti il Censimento economico che potrebbero essere oggetto di un'altra riunione da fissare entro breve tempo.