

Big data, Open data, Statistical data
Dalla qualità dei dati alla qualità dell'informazione statistica
Convegno USCI – 25-26 settembre 2014

Verso il Censimento Permanente

I comuni e il Censimento Continuo
(....si inizia dalle rilevazioni pilota e sperimentali)

N. Cecconi, N. Ferrante, M. Picci, D. Zindato

LA NUOVA STRATEGIA CENSUARIA: PUNTI CARDINE

Uso integrato di indagini campionarie e di informazioni di fonte amministrativa

- Test statistico sull'errore di copertura delle anagrafi (indagine **C-Sample**) ⇒ se l'errore di copertura è superiore alla soglia, operazioni di revisione delle anagrafi, guidate dai segnali del **SIM** (Sistema Integrato di Microdati)
- Indagine campionaria a rotazione (delle famiglie) **D-Sample** per la produzione dei dati socio-economici richiesti dal regolamento europeo e dalle esigenze nazionali
- **Rete di rilevazione *permanente***
- Uso di diverse tecniche di rilevazione *paperless*

LA NUOVA STRATEGIA CENSUARIA: PUNTI CARDINE

C-SAMPLE: Test statistico sull'errore di copertura delle anagrafi

- Campione di sezioni di censimento
- Il totale dei comuni svolgerà la rilevazione nell'arco di cinque anni. Se le anagrafi risulteranno di elevata qualità e verrà mantenuta fino alla fine del quinquennio, ogni comune sarà coinvolto una volta ogni cinque anni

D-SAMPLE: Produzione dei dati socio-economici

- Campione di famiglie
- Tutti i comuni svolgeranno la rilevazione ogni anno su un campione di famiglie senza sovrapposizione delle stesse nell'arco dei cinque anni

Verrà evitata la sovrapposizione di famiglie anche tra le due rilevazioni

INDAGINI PILOTA 2015 E SPERIMENTALI 2016

Data la complessità della nuova strategia, e al fine di limitare l'impatto sui comuni coinvolti, verranno condotte due edizioni di rilevazioni sperimentali.

- Nel **2015** verranno testate le principali innovazioni metodologiche, tecniche e organizzative (**C-Sample e D-Sample**)
- Nel **2016** (sulla base anche delle risultanze del 2015) si metteranno a punto gli aspetti metodologici, tecnici e organizzativi nei comuni >35.000 abitanti o capoluogo di provincia (comuni a maggior rischio di errori di copertura anagrafica)
 - ✓ con riferimento alla **C-Sample**, si anticiperanno le operazioni censuarie e si comincerà a testare il sistema di controllo e rientro dalla sovra/sotto-copertura
 - ✓ con riferimento alla **D-Sample**, verrà avviato il ciclo pluriennale di produzione delle stime

GENNAIO - MAGGIO 2015

- L'indagine **C-Sample** coinvolgerà tutte le regioni italiane, per un totale di **136** comuni e circa **69.000** famiglie
- L'indagine **D-Sample** coinvolgerà tutte le regioni italiane, per un totale di **151** comuni e circa **81.000** famiglie
- I comuni della C-Sample svolgeranno anche la D-Sample (nella D-sample sono stati aggiunti 15 comuni con una bassa risposta spontanea al Censimento del 2011)
- **3** strategie di rilevazione nella C-Sample e **4** nella D-Sample:
 - ✓ Le strategie della C-Sample si differenziano per i tempi della rilevazione, diversi a seconda della dimensione demografica
 - ✓ Le strategie della D-Sample si differenziano per il diverso *mix* di tecniche di rilevazione

OBIETTIVI C-SAMPLE 2015 (Gennaio – Aprile)

- Testare il sistema di produzione delle stime e il potere predittivo del SIM
- Testare l'impianto tecnologico a sostegno della rilevazione e il sistema di gestione delle operazioni sul campo (SGR_P)
- Testare l'organizzazione della rete di rilevazione
- Testare la tecnica di rilevazione e la congruità della tempistica ipotizzata
- Testare gli strumenti di comunicazione

L'INDAGINE C-SAMPLE 2015

Principali fasi della rilevazione:

- **rilevazione sul campo** *tradizionale* con tecnica CAPI che comprenderà anche la rilevazione di controllo ANNCSU
- **ritorno sul campo** con:
 - ✓ lista degli individui presenti in LAC e non trovati nella rilevazione sul campo (stimare l'errore di sovra-copertura della LAC)
 - ✓ lista degli individui presenti in LAC e non trovati ma con forti segnali di presenza in SIM
 - ✓ lista degli individui presenti in SIM e non presenti in LAC e non trovati nel corso della rilevazione

(salvaguardia della possibilità di distinguere il potere predittivo del SIM da quello della LAC)

IL QUESTIONARIO C-SAMPLE 2015

Il modello di rilevazione conterrà tre liste:

1. Lista delle ***Persone che vivevano stabilmente nell'alloggio alla data di riferimento della LAC e alla data della rilevazione vivono stabilmente nell'alloggio***
2. Lista delle ***Persone che vivevano stabilmente nell'alloggio alla data di riferimento della LAC e alla data della rilevazione non vivono più stabilmente nell'alloggio***
3. Lista delle ***Persone che non vivevano stabilmente nell'alloggio alla data di riferimento della LAC e alla data della rilevazione vivono stabilmente nell'alloggio***

Per ciascuno degli individui presenti in una delle liste verranno rilevate le variabili necessarie per effettuare il *linkage* con la LAC.

OBIETTIVI D-SAMPLE 2015 (Marzo – Maggio)

- Testare l'impianto tecnologico a sostegno della rilevazione (adeguatezza e robustezza, misure di sicurezza per la protezione dei dati sui dispositivi mobili, prestazioni generali del sistema)
- Testare il sistema di gestione delle operazioni sul campo (SGR_P)

- Testare la sostenibilità della strategia di rilevazione *paperless*
- Testare l'utilizzo della tecnica di rilevazione mista al fine del dimensionamento dei canali di restituzione
- Testare l'organizzazione della rete di rilevazione (UCC e UURR) e la congruità della tempistica ipotizzata
- Testare l'efficacia degli strumenti di comunicazione

L'INDAGINE D-SAMPLE (Marzo – Maggio 2015)

Punti cardine:

- ❖ tecnica di indagine mista (*mixed mode* simultaneo e sequenziale: auto-compilazione + intervista)
- ❖ rilevazione “*totally paperless*”
 - **utilizzo del questionario cartaceo** (file pdf stampato a cura dell'UCC) **solo in casi eccezionali** (nel caso in cui non sia possibile effettuare la compilazione elettronica del questionario)
 - **nessun questionario cartaceo inviato a Istat** ⇒ nel caso di utilizzo di questionario cartaceo, l'UCC effettuerà il data-entry utilizzando l'applicazione *web* predisposta dall'Istat
- ❖ strategie alternative di rilevazione a confronto ⇒ **differente mix di tecniche di rilevazione**

LE FASI DELLA RILEVAZIONE D-SAMPLE 2015

1- Restituzione spontanea multicanale

- **CAWI:** le famiglie potranno compilare autonomamente il questionario collegandosi al portale della rilevazione con le credenziali fornite dall'Istat
- **CAWI presso i Centri Comunali di Rilevazione (CCR):** le famiglie che non dispongono di collegamento a Internet potranno compilare il questionario on line presso i CCR
- **CATI-NUMERO VERDE:** le famiglie potranno chiamare il numero verde e compilare il questionario tramite intervista telefonica

2 - Recupero mancate risposte

- **CATI:** le famiglie saranno contattate telefonicamente dall'UCC per compilare il questionario (utenze telefoniche fornite dall'Istat + eventuali archivi comunali)
- **CAPI:** le famiglie saranno contattate a domicilio dal rilevatore (**PAPI:** canale residuale)

INFORMAZIONE E ASSISTENZA AI RISPONDENTI

- ❖ Lettera informativa inviata alle famiglie
- ❖ Durante la fase di restituzione spontanea saranno effettuati uno o più solleciti (a seconda della strategia) alle famiglie da parte dell'Istat
- ❖ I comuni potranno comunque procedere a effettuare autonomamente solleciti con gli strumenti ritenuti più idonei (lettera, telefonata, ecc.), al fine di minimizzare il ricorso ai rilevatori
- ❖ Il numero verde sarà istituito a cura di Istat. I comuni potranno comunque decidere di attivare un proprio numero verde, previa comunicazione all'ISTAT

IL QUESTIONARIO D-SAMPLE 2015

- ❖ Il Foglio di famiglia (in fase di progettazione) avrà una struttura simile al Long Form del Censimento 2011.
- ❖ La definizione dei contenuti informativi non potrà prescindere dal Regolamento Quadro 763/2008 dell'UE e dai successivi Regolamenti di attuazione predisposti per il round dei censimenti del 2010. Inoltre, si dovrà stabilire
 - ✓ se riproporre alcuni quesiti, alla luce dei risultati delle analisi effettuate sui dati del Censimento 2011
 - ✓ se inserire nuovi quesiti su tematiche emergenti
 - ✓ quali domande inserire in relazione alla misura del benessere
- ❖ Conclusa la rilevazione, si confronterà la qualità dei dati relativamente alle variabili presenti sia nel questionario che nel SIM, al fine di valutare la possibilità di sostituire parte dei dati rilevati sul campo con dati di fonte amministrativa (riduzione del carico sui rispondenti).

LA RETE DI RILEVAZIONE PERMANENTE

- La rete di rilevazione è incentrata sugli Uffici di Statistica/Servizi Demografici comunali, supportati dagli Uffici Territoriali (UURR) e dal Dipartimento per i Censimenti e gli Archivi amministrativi e statistici (DICA) dell'Istat
- Presso ogni comune dovrà essere garantito lo svolgimento delle seguenti funzioni
 - ✓ **Responsabile della rilevazione** ⇒ coordinerà e organizzerà l'attività nel comune e sarà il referente del comune per gli UURR e il DICA
 - ✓ **Referente informatico** ⇒ che dovrà essere il tramite tra il comune e l'Istat per tutte le eventuali problematiche di carattere informatico; in particolare per la C-Sample dovrà essere il riferimento per i rilevatori
 - ✓ **Operatore di *back office*** ⇒ personale preposto alle attività di coordinamento, monitoraggio delle attività sul campo, assistenza ai rispondenti (**CCR** nella D-Sample)
 - ✓ **Rilevatore** ⇒ rilevazione sul campo con modalità CAPI (intervista con dispositivo portatile)

LA CENTRALITA' DEI COMUNI

C-SAMPLE

- Coinvolti direttamente perché primi fruitori e utilizzatori dei risultati
- Coinvolti direttamente perché la “conoscenza” del territorio è un patrimonio che nessun altro soggetto può avere (società esterna)
- Coinvolti direttamente perché un maggior impegno nella “cura” dell’anagrafe, a partire dalla C-Sample, comporterà minore impegno da destinare alla rilevazione sul campo
- Coinvolti direttamente perché tra gli obiettivi che il Censimento Continuo si pone, c’è quello di un più tempestivo ed efficace supporto che la statistica può offrire alla *governance*

LA CENTRALITA' DEI COMUNI

D-SAMPLE

- Coinvolti direttamente perché primi fruitori e utilizzatori dei risultati
- Coinvolti direttamente perché occorre lavorare insieme affinché le famiglie si rendano quanto più autonome possibile rispetto agli adempimenti censuari (aumentare il tasso di risposta spontanea)
 - ✓ *meno rilevatori impegnati direttamente sul territorio*
 - ✓ *riduzione del disturbo statistico*
 - ✓ *riduzione dei costi*
- Coinvolti direttamente perché tra gli obiettivi che il Censimento Continuo si pone, c'è quello di un più tempestivo ed efficace supporto che la statistica può offrire alla *governance*